Poisson Hypothesis for Information Networks (A study in non-linear Markov processes) I. Domain of Validity

Alexander Rybko
Institute for the Information Transmission Problems,
Russian Academy of Sciences, Moscow, Russia,
rybko@iitp.ru

Senya Shlosman
Centre de Physique Theorique, CNRS,
Luminy, 13288 Marseille, France
shlosman@cpt.univ-mrs.fr

February 1, 2008

Abstract

In this paper we study the Poisson Hypothesis, which is a device to analyze approximately the behavior of large queueing networks. We prove it in some simple limiting cases. We show in particular that the corresponding dynamical system, defined by the non-linear Markov process, has a line of fixed points which are global attractors. To do this we derive the corresponding non-linear equation and we explore its self-averaging properties. We also argue that in cases of havy-tail service times the PH can be violated.

MSC-class: 82C20 (Primary), 60J25 (Secondary)

1 Introduction

The Poisson Hypothesis deals with large queueing systems. For general systems one can not compute exactly the quantities of interest, so various ap-

proximations are used in practice. The Poisson Hypothesis was formulated first by L. Kleinrock in [K]. It is the statement that certain approximation becomes exact in the appropriate limit. It concerns the following situation. Suppose we have a large network of servers, through which many customers are travelling, being served at different nodes of the network. If the node is busy, the customers wait in the queue. Customers are entering into the systems via some nods, and the external flows of customers from the outside are Poissonian. The service time at each node is random, with some fixed distribution, depending on the node. We are interested in the stationary distribution $\pi_{\mathcal{N}}$ at a given node $\mathcal{N}$: what is the distribution of the queue at it, what is the average waiting time, etc. Except for a very few special cases, when the service times are exponential, the distributions $\pi_{\mathcal{N}}$ in general can not be computed. The recipe of the Poisson Hypothesis for approximate computation of $\pi_{\mathcal{N}}$ is the following:

- consider the total flow $\mathcal{F}$ of customers to the node $\mathcal{N}$. (In general, $\mathcal{F}$ is not Poissonian, of course.) Replace $\mathcal{F}$ with a constant rate Poisson flow $\mathcal{P}$, the rate being equal to the average rate of $\mathcal{F}$. Compute the stationary distribution $\hat{\pi}_{\mathcal{N}}$ at $\mathcal{N}$, corresponding to the inflow $\mathcal{P}$. (These computations are the subject of classical queueing theory and usually provide explicit formulas.) The claim is that $\hat{\pi}_{\mathcal{N}} \approx \pi_{\mathcal{N}}$.

The Poisson Hypothesis is supposed to give a good estimate if the internal flow to every node $\mathcal{N}$ is a sum of flows from many other nodes, and each of these flows constitute only a small fraction of the total flow to $\mathcal{N}$.

Clearly, the Poisson Hypothesis can not be literally true. It can hopefully hold only after some kind of "thermodynamic" limit is taken. Its meaning is that in the long run the different nodes become virtually independent, i.e. propagation of chaos takes place. The reason for that should be that any synchronization of the nodes, if initially present, dissolves with time, due to the randomness of the service times.

In the present paper we prove the Poisson Hypothesis for the information networks in some simple cases. Namely, we will consider the following closed queueing network. Let there be M servers and N customers to be served. The distribution of the service time is given by some fixed random variable η. Upon being served, the customer chooses one of M servers with probability $\frac{1}{M}$, and goes for the service there. If there is a queue, he waits for his turn. Then in the limit $M, N \to \infty$, with $\frac{N}{M} \to \rho$, the Poisson Hypothesis holds,

under certain general restrictions on η. More precisely, **Poisson Hypothesis** *for our model* states that:

1. every k-tuple of servers becomes asymptotically mutually independent and identically distributed, as $M, N \to \infty$, for any k,

2. the total flow $\mathcal{F}_{M,N}$ of customers to a given node goes, as $M, N \to \infty$, to a Poisson flow $\mathcal{P}$,

3. the rate function $\lambda(t)$ of the Poisson flow $\mathcal{P}$ goes to a constant limit $c(\rho)$, as $t \to \infty$, which depends only on the load ρ and thus can be (easily) computed apriori.

   To be precise, one needs also to put conditions on convergence of the sequence of initial states $\nu_{M,N}$ of our servers, in order the above to hold, see Section 3 for more details.

In this paper we prove the Poisson Hypothesis (PH) in the following cases:

- for a special class of the service times η (with some exponential moments finite; actually a bit more is needed, see (12)) – with no extra conditions;

- for general η-s with heavy tails – provided the initial state of our system possesses certain desynchronization property (valid once the expected service time is finite, see (15)).

In a subsequent paper [RS] we will show that for η-s with only polynomial moments and for certain initial states the initial synchronization of the nodes may not vanish with time. So PH can be violated, we can not predict the long time behavior of a single server, and we have the phenomenon of phase transition, which manifests itself in the strong dependence on the initial state of the system.

An important step in establishing the validity of PH was made in the paper [KR1]. Namely, the properties 1 and 2 were obtained there. However, the technique of [KR1] was not enough to prove the relaxation property $\lambda(t) \to c$, and moreover it does not hold in general. It was proven there that the situation at a given single server is described by the so-called non-linear Markov process μ_t with Poissonian input with rate $\lambda(t)$, and the (non-Poissonian) output with the same rate $\lambda(t)$. The remaining problem can be formulated

as follows: this non-linear Markov process defines some complicated dynamical system, and the question is about its invariant measures. Namely, this system has one parameter family of fixed points, and one needs to show that every trajectory converges to one of them.

In the present paper we complete the program, showing that the above relaxation $\lambda(t) \to c$ indeed takes place for certain class of the service times η and certain class of initial states, and so $\mu_t \to \mu_c$, where μ_c is the stationary distribution of the stationary Markov process with the Poisson input, corresponding to constant rate $\lambda(t) = c$. In the language of dynamical systems, we find conditions under which there are no other invariant measures except these defined by the fixed points.

The central discovery of the present paper, which seems to be the key to the solution of the problem, is that, roughly speaking, the function $\lambda(t)$ has to satisfy the following non-linear equation:

$$\lambda(t) = [\lambda(\cdot) * q_{\lambda,t}(\cdot)](t). \tag{1}$$

Here $*$ stays for convolution: for two functions $a(\cdot), b(\cdot)$ it is defined as

$$[a(\cdot) * b(\cdot)](t) = \int a(t-x)\, b(x)\, dx,$$

while $q_{\lambda,t}(\cdot)$ is a family of probability densities with t real, which depends also in an implicit way on the unknown function $\lambda(\cdot)$. We call (1) the self-averaging property. The present paper consists therefore of two parts: we prove that indeed the self-averaging relation holds, and we prove then that in certain cases it implies relaxation. The self-averaging property is a special case of a more general averaging relation (29), which relates the output flow rate to the input flow rate in a queuing process with one server (called $M(t)/GI/1$ in queueing theory jargon). It appears to be new.

It is amazing that the relation (1) depends crucially on the validity of some purely combinatorial statement concerning certain problem of the placement of the rods on the line $\mathbb{R}^1$, see Section 6. This validity seems not at all obvious or to be expected, so to have it is our luck.

To fix the terminology, we remind the reader here what we mean by the **non-linear Markov process** (see [M1], [M2]). We do this for the simplest case of discrete time Markov chains, taking values in a finite set S, $|S| = k$. In such a case the set of states of this Markov chain is a simplex Δ_k of all probability measures on S, $\Delta_k = \{\mu = (p_1, ..., p_k) : p_i \geq 0, p_1 + ... + p_k = 1\}$,

while the Markov evolution defines a map $P : \Delta_k \to \Delta_k$. In the case of usual Markov chain P is affine, and this is why we will call it *linear* chain. In this case the matrix of transition probabilities coincides with P. The non-linear Markov chain is defined by a family of transition probability matrices P_μ, $\mu \in \Delta_k$, so that matrix element $P_\mu(i,j)$ is a probability of going from i to j in one step, starting *in the state* μ. The *(non-linear)* map P is then defined by $P(\mu) = \mu P_\mu$.

The ergodic properties of the linear Markov chains are settled by the Perron-Frobenius theorem. In particular, if the linear map P is such that the image $P(\Delta_k)$ belongs to the interior $\mathrm{Int}(\Delta_k)$ of Δ_k, then there is precisely one point $\mu \in \mathrm{Int}(\Delta_k)$, such that $P(\mu) = \mu$, and for every $\nu \in \Delta_k$ we have the convergence $P^n(\nu) \to \mu$ as $n \to \infty$.

In case P is non-linear, we are dealing with more or less arbitrary dynamical system on Δ_k, and the question about stationary states of the chain or about measures on Δ_k invariant under P can not be settled in general.

Therefore it is natural to ask about the specific features of our dynamical system, which permit us to find all its invariant measures. We explain this in the following subsection. The reader who is not interested in this aspect of the problem can safely skip it.

Dynamical systems aspect. Here we will use the notation of the paper, though in fact the situation of the paper is more complicated; in particular the underlying space is not a manifold, but a space of all probability measures over some non-compact set.

Let M be a manifold, supplied with the following structures:

- for every point $\mu \in M$ and every $\lambda > 0$ a tangent vector $X(\mu,\lambda)$ at μ is defined,
- a function $b : M \to R^+$ is fixed.

We want to study the dynamical system

$$\frac{d}{dt}\mu(t) = X(\mu(t), b(\mu(t))). \tag{2}$$

Its flow conserves another given function, $N : M \to R^+$, and we want to prove that our dynamical system has one-parameter family of fixed points - each corresponding to one value of N - and no other invariant measures.

We have the following extra properties of our dynamical system:

Let $\lambda(t) > 0$; consider the differential equation

$$\frac{d}{dt}\mu(t) = X(\mu(t), \lambda(t)), \ t \geq 0, \tag{3}$$

with $\mu(0) = \nu$. We denote the solution to it by $\mu_{\nu,\lambda(\cdot)}(t)$. We know that

for every $c > 0$ and every initial data ν, the solution $\mu_{\nu,\lambda(\cdot)}(t)$ to (3) converges to some stationary point $\nu_c \in M$,

$$\mu_{\nu,\lambda(\cdot)}(t) \to \nu_c, \text{provided } \lambda(t) \to c \text{as } t \to \infty, \tag{4}$$

- for the function N we have

$$\frac{d}{dt}N\left(\mu_{\nu,\lambda(\cdot)}(t)\right) = \lambda(t) - b\left(\mu_{\nu,\lambda(\cdot)}(t)\right).$$

  In particular, for every trajectory $\hat{\mu}_\nu(t)$ of (2) (where $\hat{\mu}_\nu(0) = \nu$) we have $N(\hat{\mu}_\nu(t)) = N(\nu)$. Also, $N(\nu_c)$ is continuous and increasing in c;

- for every $\nu, \lambda(\cdot)$ and every $t > 0$ there exists a probability density $q_{\nu,\lambda,t}(x), \ x \geq 0$, such that

$$b\left(\mu_{\nu,\lambda(\cdot)}(t)\right) = (\lambda * q_{\nu,\lambda,t})(t),$$

  where

$$(\lambda * q_{\nu,\lambda,t})(y) = \int_{x \geq 0} q_{\nu,\lambda,t}(x)\, \lambda(y - x)\ dx.$$

  The family $q_{\nu,\lambda,t}(x)$ satisfies:

$$\int_0^1 q_{\nu,\lambda,t}(x)\ dx = 1 \text{ for all } \nu, \lambda, t,$$

  and

$$\inf_{\substack{\nu,\lambda,t \\ x \in [0,1]}} q_{\nu,\lambda,t}(x) > 0$$

  (absolute continuity with respect to Lebesgue).

Then for every initial state ν

$$\hat{\mu}_\nu(t) \to \nu_c, \tag{5}$$

where c satisfies $N(\nu_c) = N(\nu)$.

Our statement follows from the fact that the self-averaging property,

$$f(t) = (f * q_t)(t),$$

with $q_t(\cdot)$ being a family of probability densities on $[0,1]$, implies that $f(t) \to const$ as $t \to \infty$, so (5) follows from (4). This implication is the subject of the Theorems 21, 22, 25.

We feel that the relation (1) is an important feature of the subject we are interested in. Therefore in the present paper we study it and the related questions in some generality.

i) We start with the equation

$$f(t) = [f(\cdot) * q_t(\cdot)](t). \tag{6}$$

Here we suppose that $q_t(\cdot)$ is just some one-parameter family of probability densities (independent of f), so (1) is a special case of (6). On the other hand, we suppose additionally that all the distributions $q_t(\cdot)$ are supported by some finite interval. We establish relaxation in this case.

ii) We then do the same for the case of distributions $q_t(\cdot)$ with unbounded support.

iii) Last, we treat the true problem, where in addition to the infinite support, an extra parameter μ appears and an extra perturbation is added to convolution term in (6) :

$$\lambda(t) = (1 - \varepsilon_{\lambda,\mu}(t)) [\lambda(\cdot) * q_{\lambda,\mu,t}(\cdot)](t) + \varepsilon_{\lambda,\mu}(t) Q_{\lambda,\mu}(t). \tag{7}$$

Here the parameter $\varepsilon_{\lambda,\mu}(t)$ is small: $\varepsilon_{\lambda,\mu}(t) \to 0$ as $t \to \infty$, the term $Q_{\lambda,\mu}(t)$ is uniformly bounded, while the meaning of μ will be explained later.

As we proceed from *i*) to *iii*), we will have to assume more about the class of distributions $\{q_.\}$, for which the self-averaging implies relaxation.

We finish this introduction by a brief discussion of the previous work on the subject, and their methods.

As we said before, part of the proof of the Poissonian Hypothesis – the so called Weak Poissonian Hypothesis – was obtained in [KR1]. By proving that the Markov semigroups describing the Markov processes for finite M, N, after factorization by the symmetry group of the model converge, as $M, N \to \infty$, $\frac{N}{M} \to \rho$, to the semigroup, describing the non-linear Markov process, the authors have proven that the limit flows to each node are independent Poisson flows with the same rate function $\lambda(t)$. This statement is

fairly general, and can be generalized to other models with the same kind of the symmetry – the so-called mean-field models. The general theory – see, for example, [L] – implies then, that all the limit points of the stationary measures of the Markov processes with finite M, N are invariant measures of the limiting non-linear Markov process. The remaining step – the proof that the limiting dynamical system has no other attractors except the one-parameter family of the fixed points – is done in the present paper *for some class of the service times* η. Apriori this fact is not at all clear, and one can construct natural examples of the systems with many complicated attractors, which are reflected in the complex behavior of the Markov processes with finite M, N. However, the self-averaging property, explained above, rules out such a possibility. It seems that the self-averaging property can also be generalized to other mean-field models. In a forthcoming paper [RS] we show that for more general service times the corresponding dynamical system has more complicated attractors, so that the relaxation $\lambda(t) \to c$ might or might not hold, depending on the initial state.

The Poisson Hypothesis was fully established in a pioneer paper [St] for a special case when the service time is non-random. This is a much simpler case, and the methods of the paper can not be extended to our situation. They are sufficient for a simpler case of the Poissonian service times, which case was studied in [KR2].

The paper [DKV] deals with another mean-field model, describing some open queueing network. Though the Poisson Hypothesis does not hold for it, the spirit of the main statement there is the same as in the present paper: the limiting dynamical system has precisely one global attractor, which corresponds to the fixed point.

One of specific feature of the method of the paper [DKV], as well as related paper [DF], is that the Markov processes have countable sets of values. They also correspond to open networks, when the customers are coming into the system from the outside and leave it after being served. So one can in principle use monotonicity arguments and stochastic domination. In our situation the phase space is (one-dimensional) real manifold, while the number of customers is fixed, and this technique does not seem to be applicable.

The importance of the Poisson Hypothesis as the central problem of the theory of large queueing systems was emphasized, among others, by Roland Dobrushin [D2] and Alexander Borovkov [B].

The organization of the paper is the following:

In the next Section 2 we introduce notation used in the rest of the paper. We formulate the properties needed of the distribution of the service time η – the main parameter of our model.

In the Section 3 we recall more results of the paper [KR1].

In the next Section 4 we formulate out main result.

In Section 5 we derive the self-averaging relation (1), our key tool in establishing the Poisson Hypothesis. This is a statement about single server queue, $M(t)/GI/1$, in queueing theory jargon. This and the next sections are self-sufficient, and there we do not need any condition at all on the service time distribution.

The Section 6 contains the proof of a combinatorial statement dealing with the rod placements on $\mathbb{R}^1$. It is a key statement used in the previous section. It seems also to be of independent interest.

In the next Section 7 we have collected various technical statements and estimates used in the sequel. Again, some of them, – for example, the calculus Lemma 12 – seem to be of independent interest.

The Section 8 contains the derivation of the noisy version of the self-averaging relation (1) – the relation (26). Again, it uses the combinatorial statement of the Section 6.

In Section 9 we consider a simplified case of our theory, when there are infinitely many servers at each node, so there are no queues. This case is presented for pedagogical reasons only. It can be analyzed by application of renewal theory and proving the Local Limit Theorem for i.i.d. random variables.

In the next Section 10 we explain that in more realistic situation one needs the Local Limit Theorem for Markov chain, instead of i.i.d.-s, but we show that it does not hold in general, so the analog of the renewal theory needed there does not exist as well. This is why the rest of the paper uses the analytic methods quite a bit.

The next three Sections 11, 12 and 13 contain the main step of our proof: the derivation of the relaxation property from the self-averaging relation. The Section 11 deals with the case of finite range averaging kernels, the Section 12 – with the infinite range kernels, while the Section 13 – with the noisy case. As we proceed from easier cases to more difficult, the generality of our theorems becomes less and less.

The last Section 14 contains some conclusions and lists possible directions of further research.

The reader who wants to go directly to the proof of the general result, can jump from Section 8 straight to Section 12.1 and then to Section 13.

2 Notation

In this section we will fix the notation for the non-linear Markov process, which describes a given server in the above described limit.

Server. It is defined by specifying the distribution of the random serving time η, i.e. by the function

$$F(t) = \Pr\{\text{serving time } \eta \le t\}.$$

We suppose that η is such that:

1. the density function $p(t)$ of η is positive on $t \ge 0$ and uniformly bounded from above;

2. $p(t)$ satisfies the following strong Lipschitz condition: for some $C < \infty$ and for all $t \ge 0$
$$|p(t+\Delta t) - p(t)| \le Cp(t)|\Delta t|, \tag{8}$$
provided $t + \Delta t > 0$ and $|\Delta t| < 1$;

3. introducing the random variables
$$\eta\Big|_{\tau} = \left(\eta - \tau \,\Big|\, \eta > \tau\right), \tau \ge 0,$$
we suppose that for some $\delta > 0$, $M_{\delta,\tau} < \infty$
$$\mathbb{E}\left(\eta\Big|_{\tau}\right)^{2+\delta} < M_{\delta,\tau}. \tag{9}$$
Of course, this condition holds once
$$M_{\delta} \equiv \mathbb{E}(\eta)^{2+\delta} < \infty. \tag{10}$$
In what follows, the function
$$R_{\eta}(\tau) \equiv \mathbb{E}\left(\eta\Big|_{\tau}\right) < \infty \tag{11}$$

will play a crucial role. In particular, if

$$R_\eta(\tau) < \bar{C} \tag{12}$$

for all $\tau \geq 0$, then PH holds for every initial state, as we will explain later. However, the condition (12) is too restrictive; it implies that the random variable η has some exponential moments finite (though the opposite is not true). *Generally we are not assuming* (12).

4. Without loss of generality we can suppose that

$$\mathbb{E}(\eta) = 1. \tag{13}$$

In what follows, the function $p(\cdot)$ will be fixed.

The remaining two conditions will not be used explicitly in the present paper. However we have to impose them since they are used in the paper [KR1], while we are using its results:

5. the probability density $p(t)$ is differentiable in t, with $p'(t)$ continuous. Moreover, introducing the functions $p_\tau(t)$ as the densities of the random variables $\eta\Big|_\tau$, we require that the function $p_\tau(0)$ is bounded uniformly in $\tau \geq 0$,

$$p_\tau(0) \leq U(\eta), \tag{14}$$

while the function $\frac{d}{d\tau}p_\tau(0)$ is continuous and bounded uniformly in $\tau \geq 0$;

6. the limits $\lim_{\tau\to\infty} p_\tau(0), \lim_{\tau\to\infty} \frac{d}{d\tau}p_\tau(0)$ exist and are finite.

Configurations. By a configuration of a server at a given time moment t we mean the following data:

- The number $n \geq 0$ of customers waiting to be served. The customer who is served at t, is included in the total amount n. This quantity n will be called *the length of the queue.*

- The duration τ of the elapsed service time of the customer under the service at the moment t.

Therefore the set of all configurations Ω is the set of all pairs (n, τ), with an integer $n > 0$ and a real $\tau > 0$, plus the point $\mathbf{0}$, describing the situation of the server being idle. For a configuration $\omega = (n, \tau) \in \Omega$ we define $N(\omega) = n$. We put $N(\mathbf{0}) = 0$.

States. By a state of the system we mean a probability measure μ on Ω. We denote by $\mathcal{M}(\Omega)$ the set of all states on Ω.

Observables. There are some natural random variables associated with our system. One is the queue length in the state μ, $N_\mu = N_\mu(\omega)$. We denote by $N(\mu)$ the mean queue length in the state μ :

$$N(\mu) = \mathbb{E}(N_\mu) \equiv \langle N_\mu(\omega) \rangle_\mu ,$$

and we introduce the subsets $\mathcal{M}_q(\Omega) \subset \mathcal{M}(\Omega)$, $q \geq 0$ by

$$\mathcal{M}_q(\Omega) = \{\mu \in \mathcal{M}(\Omega) : N(\mu) = q\} .$$

Another one is the expected service time S_μ, corresponding to the function

$$S(\omega) = \begin{cases} 0 & \text{for } \omega = \mathbf{0}, \\ (n-1) + R_\eta(\tau) & \text{for } \omega = (n, \tau), \text{with } n > 0. \end{cases}$$

Again, we define

$$S(\mu) = \mathbb{E}(S_\mu) \equiv \langle S(\omega) \rangle_\mu . \tag{15}$$

Clearly, if the condition (12) holds, then

$$S(\mu) \leq \bar{C} N(\mu) , \tag{16}$$

and therefore $S(\mu)$ is finite once $\mu \in \mathcal{M}_q(\Omega)$ for some q. In general, the expected service time $S(\mu)$ can be infinite for some states μ. These are the states for which PH can be violated, as we explain in [RS].

Input flow. Let a function $\lambda(t) \geq 0$ is given. We suppose that the input flow to our server is a Poisson process with rate function $\lambda(t)$, which means in particular that the probabilities $P_k(t, s)$ of the events that k new customers arrive during the time interval $[t, s]$ satisfy

$$P_k(t, t + \Delta t) = \begin{cases} \lambda(t) \Delta t + o(\Delta t) & \text{for } k = 1, \\ 1 - \lambda(t) \Delta t + o(\Delta t) & \text{for } k = 0, \\ o(\Delta t) & \text{for } k > 1, \end{cases}$$

as $\Delta t \to 0$, while for non-intersecting time segments $[t_1, s_1]$, $[t_2, s_2]$ the flows are independent.

Output flow. Suppose the initial state $\nu = \mu(0)$, as well as the rate function $\lambda(t)$, with $\lambda(t) = 0$ for $t < 0$, of the input flow are given. Then the system evolves in time, and its state at the moment t is given by the measure

$$\mu(t) = \mu_{\nu,\lambda(\cdot)}(t).$$

In particular, the probabilities $Q_k(t,s) = Q_k(t,s;\nu,\lambda(\cdot),p(\cdot))$ of the events that k customers have finished their service during the time interval $[t,s]$ are defined. We suppose that the customer, once served, leaves the system.

The resulting random point process $Q_\cdot(\cdot,\cdot)$ need not, of course, be Poissonian. However we still can define its rate function $b(t)$ as the one satisfying

$$Q_k(t,t+\Delta t) = \begin{cases} b(t)\,\Delta t + o(\Delta t) & \text{for } k=1, \\ 1 - b(t)\,\Delta t + o(\Delta t) & \text{for } k=0, \\ o(\Delta t) & \text{for } k>1, \end{cases}$$

as $\Delta t \to 0$. The rate function $b(\cdot)$ of the output flow is determined once the initial state $\nu = \mu(0)$ and the rate function $\lambda(\cdot)$ of the input flow are given. Therefore the following (non-linear) operator A is well defined:

$$b(\cdot) = A(\nu,\lambda(\cdot)).$$

We will call the general situation, described by the pair $\nu,\lambda(\cdot)$, and $b(\cdot) = A(\nu,\lambda(\cdot))$, as a General Flow Process (GFP). (This time inhomogeneous (linear) Markov process is usually labelled in queueing theory by $M(t)/GI/1$.)

The following is known about the operator A, see [KR1]:

- For every initial state ν the equation

  $$A(\nu,\lambda(\cdot)) = \lambda(\cdot)$$

  – in words: *the rate of the input equals the rate of the output* – has exactly one solution $\lambda(\cdot) = \lambda_\nu(\cdot)$. Then the evolving state $\mu_{\nu,\lambda_\nu(\cdot)}(t)$ is *the non-linear Markov process,* which we will abbreviate as NMP.

- This non-linear Markov process has the following conservation property: for all t

  $$N\left(\mu_{\nu,\lambda_\nu(\cdot)}(t)\right) = N(\nu)$$

  (because "the rates of the input flow and the output flow coincide"). So the spaces $\mathcal{M}_q(\Omega)$ are invariant under non-linear Markov evolutions.

- All the functions $\lambda_\nu(\cdot)$ are bounded:

$$\lambda_\nu(t) \leq C = C(\eta) \tag{17}$$

uniformly in ν and t. (This is clear, since the output flow has its rate uniformly bounded, by the constant $U(\eta)$, see (14). So (17) holds with $C(\eta) = U(\eta)$.)

- For every constant $c \in [0,1)$ there exists the initial state ν_c, such that

$$A(\nu_c, c) = c. \tag{18}$$

(Here we identify the constant c with the function taking the value c everywhere.) Moreover, this measure ν_c is a stationary state: $\mu_{\nu_c,c}(t) = \nu_c$ for all $t > 0$. The function $c \rightsquigarrow N(\nu_c)$ is continuous increasing, with $N(\nu_0) = 0$, $N(\nu_c) \uparrow \infty$ as $c \to 1$.

The non-linear Markov process $\mu_{\nu,\lambda_\nu(\cdot)}(t)$ is the main object of the present paper. Therefore we will give now another definition of this process, via jump rates of transitions during the infinitesimal time, Δt. So suppose that our process is in the state $\mu \in \mathcal{M}(\Omega)$, and assumes the value $\omega = (n,\tau) \in \Omega$. During the time increment Δt the following two transitions can happen with probabilities of order of Δt :

- the customer under the service will finish it and will leave the server, so the value $\omega = (n,\tau)$ will become $(n-1,\varsigma)$, with $\varsigma \leq \Delta t$. The probability of this event is

$$c_1 \Delta t + o(\Delta t),$$

where

$$c_1 = c_1(\omega) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \frac{\int_\tau^{\tau+\Delta t} p(x)\, dx}{\int_\tau^\infty p(x)\, dx},$$

while for $\omega = \mathbf{0}$ we put $c_1(\mathbf{0}) = 0$;

- a new customer will arrive to the server, so the value $\omega = (n,\tau)$ will become $(n+1, \tau + \Delta t)$. The probability of this event is given by

$$c_2 \Delta t + o(\Delta t),$$

where the rate c_2 depends on the whole state μ, and does not depend on ω :

$$c_2 = c_2(\mu) = \mathbb{E}_\mu(c_1(\omega)).$$

In words, the input rate is the average rate of the output in the state μ.

It is curious to note that while for the general nonlinear continuous time Markov processes its rates depend both on the configuration and on the state of the process, in our case the rate c_1 depends only on the configuration, while the rate c_2 – only on the state of the process.

3 More facts from [KR1]

Consider the following continuous time Markov process $\mathfrak{M}$. Let there be M servers and N customers. The serving times are i.i.d., with distribution η. The configuration of the system consists of specifying the numbers of customers $n_i,\ i = 1, ..., M$, waiting at each server, plus the duration τ_i of the service time for every customer under service. Therefore it is a point in

$$\Theta_{M,N} = \left\{(\omega_1, ..., \omega_M) \in \Pi_{i=1}^M \Omega_i : n_1 + ... + n_M = N\right\}.$$

Upon being served, the customer goes to one of M servers with equal probability $1/M$, and is there the last in the queue.

The permutation group $\mathcal{S}_M$ acts on $\Theta_{M,N}$, leaving the transition probabilities invariant. Therefore we can consider the factor-process. Its values are (unordered) finite subsets of Ω. It can be equivalently described as a measure

$$\nu = \frac{1}{M}\sum_{i=1}^M \delta_{(n_i,\tau_i)}.$$

We will identify such measures with the configurations of the symmetrized factor-process. Note that

$$\langle n\rangle_\nu = \frac{N}{M}.$$

So if we use the notation $\mathcal{M}_\rho(\Omega) \subset \mathcal{M}(\Omega)$ for the measures μ on Ω for which $\langle n\rangle_\mu = \rho$, then we have that $\nu \in \mathcal{M}_{\frac{N}{M}}(\Omega)$. We also introduce the notation

$\mathcal{M}_{\frac{N}{M},M}(\Omega) \subset \mathcal{M}_{\frac{N}{M}}(\Omega)$ for the family of atomic measures, such that each atom has a weight $\frac{k}{M}$ for some integer k.

A state of our Markov process is a probability measure on the set of configurations, i.e. an element of $\mathcal{M}(\mathcal{M}(\Omega))$. If the initial state of the process is supported by $\mathcal{M}_\rho(\Omega)$, then at any positive time it is still the element of $\mathcal{M}(\mathcal{M}_\rho(\Omega))$. A natural embedding $\mathcal{M}(\Omega) \subset \mathcal{M}(\mathcal{M}(\Omega))$, which to each configuration $\nu \in \mathcal{M}(\Omega)$ corresponds the atomic measure δ_ν, will be denoted by δ.

For $\mu_0 = \delta_\nu \in \mathcal{M}\left(\mathcal{M}_{\frac{N}{M},M}(\Omega)\right)$ to be the initial state of our Markov process, we denote by μ_t the evolution of this state. (Clearly, in general $\mu_t \notin \delta(\mathcal{M}(\Omega))$ for positive t.) This process is ergodic. We denote by $\pi_{M,N}$ the stationary measure of this process.

Let now $\kappa \in \mathcal{M}_\rho(\Omega)$ be some measure, let the sequences of integers $N_j, M_j \to \infty$ be such that $\frac{N_j}{M_j} \to \rho$, and let the measures $\nu^j \in \mathcal{M}_{\frac{N_j}{M_j},M_j}(\Omega)$ be such that $\nu_j \to \kappa$ weakly. Consider the Markov processes $\mu_t^j \in \mathcal{M}\left(\mathcal{M}_{\frac{N_j}{M_j},M_j}(\Omega)\right)$, corresponding to the initial conditions δ_{ν^j}. As we just said, in general $\mu_t^j \notin \delta\left(\mathcal{M}_{\frac{N_j}{M_j},M_j}(\Omega)\right)$ for each j, once $t > 0$. However, for the limit $\mu_t = \lim_{j\to\infty} \mu_t^j$ we have that $\mu_t \in \mathcal{M}(\mathcal{M}_\rho(\Omega))$, and moreover $\mu_t \in \delta(\mathcal{M}_\rho(\Omega))$, so we can say that the random evolutions μ_t^j tend to the non-random evolution μ_t with $\mu_0 \equiv \kappa$, as $N_j, M_j \to \infty$.

Therefore we have a dynamical system

$$\mathcal{T}_t : \mathcal{M}_\rho(\Omega) \to \mathcal{M}_\rho(\Omega). \tag{19}$$

This dynamical system μ_t is nothing else but the non-linear Markov process, discussed above.

Another way of obtaining the same dynamical system is to look on the behavior of a given server. Here instead of taking the symmetrization of the initial process $\mathfrak{M}$ on $\Theta_{M,N}$, we have to consider its projection on the first coordinate, Ω_1, say. To make the correspondence with the above, we have to take for the initial state of this process a measure $\tilde{\nu}^j$ on $\Theta_{M,N}$, which is $\mathcal{S}_M$-invariant, and which symmetrization is the initial state ν^j of the preceding paragraph. The projection of $\mathfrak{M}$ on Ω_1 would not be, of course, a Markov process. However, it becomes the very same non-linear Markov process μ_t in the above limit $N_j, M_j \to \infty$.

We can generalize further, and study the projection of $\mathfrak{M}$ to a finite product, $\prod_{j=1}^{r}\Omega_j$. Then in the limit $N_j, M_j \to \infty$ this projection converges to a process on $\prod_{j=1}^{r}\Omega_j$, which factors into the product of r independent copies of the same non-linear Markov process μ_t. This statement is known as the "propagation of chaos" property.

The main result of the present paper is in particular to give conditions under which for every ρ the dynamical system (19) has exactly one fixed point ν_c, $c = c(\rho)$, and that it is globally attractive. That would imply that $\pi_{N_j,M_j} \to \nu_c$, provided $\frac{N_j}{M_j} \to \rho$ as $j \to \infty$ and $c = c(\rho)$.

4 Main result

Our main result states that under certain restrictions PH holds. In view of what was said in the beginning of the Introduction, it is sufficient to prove the following:

Theorem 1 *Consider the system, described in the Section 2, and let its service time η has properties 1-6 of this Section. For every initial state ν with finite expected service time and finite mean queue:*

$$S(\nu) < \infty, \quad N(\nu) < \infty, \tag{20}$$

the solution $\lambda_\nu(\cdot)$ of the equation

$$A(\nu, \lambda(\cdot)) = \lambda(\cdot)$$

has the ***relaxation*** *property:*

$$\lambda_\nu(t) \to c \text{ as } t \to \infty,$$

where the constant c satisfies

$$N(\nu) \equiv \mathbb{E}_\nu(N(\omega)) = N(\nu_c) \equiv \mathbb{E}_{\nu_c}(N(\omega)).$$

Moreover, $\mu_{\nu,\lambda_\nu(\cdot)}(t) \to \nu_c$ weakly, as $t \to \infty$.

In particular, if the service time η has the property that $R_\eta(\tau) < \bar{C}$ for all τ (see (12)), then the relaxation property holds for every initial state ν.

A special case of the above theorem is the following

Proposition 2 *Let $T > 0$ be some time moment, and suppose that the function $\lambda(\cdot)$ satisfies*

$$\lambda(t) = b(t) \text{ for all } t \geq T, \tag{21}$$

where

$$b(\cdot) = A(\mathbf{0}, \lambda(\cdot)). \tag{22}$$

Suppose that

$$\int_0^T \lambda(t)\, dt \leq C < \infty.$$

Then for some $c \geq 0$

$$\lambda(t) \to c \text{ as } t \to \infty. \tag{23}$$

Our theorem follows from the Proposition 2 immediately in the special case when the initial state ν is of the form $\nu = \mu_{\mathbf{0},\lambda(\cdot)}(t)$ for some λ and some $t > 0$. These initial states are easier to handle, so we treat them separately.

The heuristics behind the Proposition 2 is the following. One expects that if

$$b(\cdot) = A(\nu, \lambda(\cdot)),$$

then the function b for large times is "closer to a constant" than the function λ. More precisely, if t belongs to some segment $[T_1, T_2]$, with $T_1 \gg 1$, then the dependence of $b(t)$ on ν is very weak, so b is determined mainly by λ. One then argues that once the segment $[T_1, T_2]$ is large, $\sup_{t\in[T_1,T_2]} b(t)$ should be strictly less than $\sup_{t\in[T_1,T_2]} \lambda(t)$. Indeed, one can visualize the random configuration of the exit moments y_i-s as being obtained from the input moments configuration of x_i-s by making it *sparser*. Namely, we have to consider a sequence η_i of i.i.d. random variables, having the same distribution as η, and then to shift the particles x_i to the right, positioning them at locations z_i, so that in the result

$$z_{i+1} - z_i \geq \eta_i \tag{24}$$

for all i-s, and $y_i = z_i + \eta_i$. Note that before the shift it might have been that $x_{i+1} - x_i < \eta_i$ for some i-s, see (31), (32) below for more details. However this is a very rough idea, since some particles need not be moved, due to the fact that $x_{i+1} - x_i \geq \eta_i$ may hold already before to the sparsening step, in which case it will happen that $z_{i+1} = x_{i+1}$, while $z_i > x_i$, and so the configuration becomes locally denser. (And if λ is a constant, then b is this same constant, so again the above argument is not literally true.)

To be more precise, we will show the following **self-averaging property**. Let the functions $\lambda(\cdot)$ and $b(\cdot)$ are related by

$$b(\cdot) = A(\mathbf{0}, \lambda(\cdot)).$$

One of the main points of the following will be to show that for every value x one can find a probability density $q_{\lambda,x}(t)$, vanishing for $t \leq 0$, such that

$$b(x) = [\lambda * q_{\lambda,x}](x). \tag{25}$$

We then will show that this self-averaging property of the system implies (23), provided we know in advance certain regularity properties of the family $\{q_{\lambda,x}\}$. Note that apriori the condition (25) is not evident at all for our FIFO (=First-In-First-Out) system: one has to rule out the situation that, say, the input rate function λ is uniformly bounded from above by $1/3$, while the output rate b is occasionally reaching the level $2/3$; this is clearly inconsistent with (25).

In general case, when

$$b(\cdot) = A(\mu, \lambda(\cdot))$$

and $S(\mu) < \infty$, we have

$$b(x) = (1 - \varepsilon_{\lambda,\mu}(x))[\lambda * q_{\lambda,\mu,x}](x) + \varepsilon_{\lambda,\mu}(x) Q_{\lambda,\mu}(x), \tag{26}$$

where $\varepsilon_{\lambda,\mu}(x) > 0$, $\varepsilon_{\lambda,\mu}(x) \to 0$ as $x \to \infty$, while $Q_{\lambda,\mu}(x)$ is a bounded term, see Section 8 for details.

5 The self-averaging relation

Here we will derive a formula, expressing the function $b(\cdot) = A(\mathbf{0}, \lambda(\cdot))$ in terms of the functions $\lambda(\cdot)$ and the density $p(\cdot)$ of η. This will be the needed self-averaging relation (25).

First, we define the kernels $q_{\lambda,x}(t)$. To do it, let $e(u)$ be the probability that our server is idle at the time u. (Note that the dependence of $e(u)$ on λ is only via $\{\lambda(s), s \leq u\}$.) Now define the function $c(u,t)$ as follows. Let us condition on the event that the server is idle just before time u, while at u the customer arrives. Under this condition define

$$c(u,t) = \lim_{h \searrow 0} \frac{1}{h} \Pr \left\{ \begin{array}{c} \text{the server is never idle during } [u, u+t]; \\ \text{during } [u+t, u+t+h] \text{ the server gets} \\ \text{through with some client} \end{array} \right\}. \tag{27}$$

Then

$$q_{\lambda,x}(t) = e(x-t)\,c(x-t,t). \tag{28}$$

Theorem 3 *Let the functions $b(\cdot)$ and $\lambda(\cdot)$ are related by*

$$b(\cdot) = A(\mathbf{0}, \lambda(\cdot)).$$

Then also

$$b(x) = \int_0^\infty \lambda(x-t)\,q_{\lambda,x}(t)\;dt. \tag{29}$$

Moreover, for all λ, x

$$\int_0^\infty q_{\lambda,x}(t)\;dt \le 1. \tag{30}$$

Note. The relation (30) is not at all obvious, and we see no way to derive it from (28) without going into the details of the serving mechanism. In fact, it does not hold for some more general models.

Proof. We first introduce some new notions.

Let $l_1, ..., l_n > 0$ be a collection of positive real numbers, which we will interpret as the lengths of hard rods ($\equiv$service times), placed in $\mathbb{R}^1$. A configuration of rods can be then given by specifying the sequence $x_1, x_2, ..., x_n$ of their left-ends: the rod l_i occupies the segment $[x_i, x_i + l_i]$. This configuration will be denoted by $\sigma_n(x_1, x_2, ..., x_n; l_1, ..., l_n)$.

In case some of the rods from $\sigma_n(x_1, x_2, ..., x_n; l_1, ..., l_n)$ are intersecting over a nondegenerate segments, we say that such a configuration has conflicts. By a resolution of conflicts we call another placement of the rods $l_1, ..., l_n$ on the line. To define it, we first need to reenumerate the points of the sequence $x_1, x_2, ..., x_n$ so that it will become increasing. To save on notation we suppose *until the end of this paragraph only* that it is initially so. Then the new placing of the rods have the following sequence $z_1 < z_2 < ... < z_n$ of the left-ends:
it is defined inductively by

$$z_1 = x_1,$$

and

$$z_i = \max\{z_{i-1} + l_{i-1}, x_i\} \tag{31}$$

(Lindley equation). We will denote by y-s the corresponding set of the right-ends:

$$y_i = z_i + l_i. \tag{32}$$

Any configuration with no conflicts, and in particular any configuration obtained by resolution of the conflicting one, will be called an **r-configuration**. The operation of resolving the conflict will be denoted by R, so

$$\sigma_n\left(z_1, z_2, ..., z_n; l_1, ..., l_n\right) = R\sigma_n\left(x_1, x_2, ..., x_n; l_1, ..., l_n\right).$$

(The R-operation is not well defined for x-sequences with coinciding entries. However, they have zero Lebesgue measure, which makes them irrelevant for our future needs.)

For any configuration σ of rods we will denote by $Y(\sigma)$ the set of their right-ends. So, in our notations

$$\left(y_1, ..., y_n\right) = Y\left(R\sigma_n\left(x_1, x_2, ..., x_n; l_1, ..., l_n\right)\right).$$

Suppose now that the lengths $l_1, ..., l_n$, the set $x_1, x_2, ..., x_{n-1}$ (with $n-1$ points) and the location $y \in \mathbb{R}^1$ are specified. We define the values $X(y) \equiv X\left(y \,\middle|\, x_1, x_2, ..., x_{n-1}; l_1, ..., l_n\right) \in \mathbb{R}^1$ as the solutions of the equation

$$y \in Y\left(R\sigma_n\left(x_1, x_2, ..., x_{n-1}, X(y); l_1, ..., l_n\right)\right). \tag{33}$$

It is clear that the function $X\left(y \,\middle|\, x_1, x_2, ..., x_{n-1}; l_1, ..., l_n\right)$ is not defined everywhere, and on the set where it is defined, it is multivalued, provided $n \geq 2$. (The case $n=1$ is trivial: $X\left(y \,\middle|\, l_1\right) = y - l_1$.) However, outside the set of x-s and l-s of Lebesgue measure zero in $\mathbb{R}^{2n-1}$, which set is irrelevant for our future purposes, its multivaluedness is reduced to "finitely-many-valuedness".

Now we can write the desired formula:

$$b(y) = \exp\left\{-I_\lambda(y)\right\} \sum_{n=1}^{\infty} \frac{1}{(n-1)!} \times \tag{34}$$

$$\times \underbrace{\int_0^\infty ... \int_0^\infty}_{n} \left[\underbrace{\int_0^y ... \int_0^y}_{n-1} \lambda\left(X\left(y \,\middle|\, \left\{x_1, ..., x_{n-1}\right\}; l_1, ..., l_n\right)\right) \prod_{i=1}^{n-1} \lambda\left(x_i\right)\, dx_i\right] \prod_{i=1}^{n} p\left(l_i\right)\, dl_i,$$

where

$$I_\lambda(y) = \int_0^y \lambda(x)\, dx.$$

The integral in (34) should be understood as follows: the range of integration coincides with the domain where the function $X\left(y \,\middle|\, \{x_1, ..., x_{n-1}\}; l_1, ..., l_n\right)$ is defined, while over the domains where the function X is multivalued one should integrate each branch separately and then take the sum of integrals.

In words, the meaning of the relation (34) is the following: for every realization $x_1, ..., x_{n-1}$ of the Poisson random field and every realization $l_1, ..., l_n$ of the sequence of the service times, we look for time moments $X = X\left(y \,\middle|\, x_1, ..., x_{n-1}; l_1, ..., l_n\right)$, at which the l_n-customer has to arrive, so as to ensure that at the moment y some customer (perhaps a different one) will exit, after being served. That is, moments $X = X\left(y \,\middle|\, x_1, ..., x_{n-1}; l_1, ..., l_n\right)$ are beginnings of busy periods, during which there happens an exit at time y. In some cases such moments might not exist, while in other cases there might be more than one such moment. If X_i are these moments, we then have to add all the rate values, $\lambda(X_i)$, and to integrate the sum $\sum_i \lambda(X_i)$ over all n and all $x_1, ..., x_{n-1}; l_1, ..., l_n$, thus getting the exit rate $b(y)$. A one-second thought will convince the reader that the formula (34) contains in itself another definition of the kernel (28), together with the proof of the relation (29). So we need only to prove (30), which turns out to be quite delicate.

The first summand ($n = 1$) in the sum in (34) is by definition the convolution,

$$b_1(y) = \int_0^y \lambda(y - l)\, p(l)\, dl. \tag{35}$$

Since $p(l) \geq 0$ and

$$\int_0^y p(l)\, dl \leq 1, \tag{36}$$

we have indeed that $b_1(y) < \sup_{x \leq y} \lambda(x)$ in case when, say, the maxima of λ are isolated, or when λ is not a constant and the support of the distribution p is the full semiaxis $\{l > 0\}$. We want to show that in some sense the same is true for all the functions b_n, defined as

$$b_n(y) = \int \left[\int \lambda\left(X\left(y \,\middle|\, x_1, ..., x_{n-1}; l_1, ..., l_n\right)\right) \prod_{i=1}^{n-1} \left(\frac{\lambda(x_i)}{I_\lambda(y)}\, dx_i \right) \right] \prod_{i=1}^{n} p(l_i)\, dl_i. \tag{37}$$

Since

$$b(y) = \exp\{-I_\lambda(y)\} \sum_{n=1}^{\infty} \frac{I_\lambda(y)^{n-1}}{(n-1)!} b_n(y),$$

the crucial step will be the analog of (35), (36) for all $n > 1$, that is that

$$b_n(y) = \int_0^y \lambda(y-l)\, p_n(l)\ dl,$$

for some $p_n(l) \geq 0$, $\int_0^y p_n(l)\ dl \nearrow 1$ for $y \to \infty$. This turns out to be quite an involved combinatorial statement.

Note that, evidently, the measure $\prod_{i=1}^n p(l_i)\ dl_i$ is invariant under the coordinate permutations in $\mathbb{R}^n$; therefore we can rewrite the expression (37) for the function $b_n(y)$ as

$$b_n(y) = \int \left[\int \frac{1}{n!} \lambda\left(\bar{X}\left(y \,\middle|\, x_1, ..., x_{n-1}; \{l_1, ..., l_n\} \right) \right) \prod_{i=1}^{n-1} \left(\frac{\lambda(x_i)}{I_\lambda(y)}\, dx_i \right) \right] \prod_{i=1}^{n} p(l_i)\ dl_i, \tag{38}$$

where the following notations and conventions are used:

- the (multivalued) function $\bar{X}\left(y \,\middle|\, x_1, ..., x_{n-1}; \{l_1, ..., l_n\} \right)$ by definition assigns to every y the union of the sets of solutions $X(y)$ of all the equations

$$y \in Y\left(R\sigma_n\left(x_1, ..., x_{n-1}, X(y); l_{\pi(1)}, ..., l_{\pi(n)} \right) \right), \tag{39}$$

  with π running over all the permutation group $\mathcal{S}_n$ (the notation $\{l_1, ..., l_n\}$ stresses the fact that the function $\bar{X}$ does not depend on the order of l_i-s);

- the entries of the set $\bar{X}\left(y \,\middle|\, x_1, ..., x_{n-1}; \{l_1, ..., l_n\} \right)$ have to be counted with multiplicities, which for a given $x \in \bar{X}\left(y \,\middle|\, x_1, ..., x_{n-1}; \{l_1, ..., l_n\} \right)$ is by definition the number of equations (39) with different π-s, to which x is a solution;

- the integration in (38) of the multivalued function means that each sheet should be integrated and the results added. Moreover, each sheet has to be taken as many times as its multiplicity is.

Since each contribution $\lambda\left(X\left(y \,\middle|\, x_1, ..., x_{n-1}; l_1, ..., l_n \right) \right)$ to (37) appears $n!$ times in (38), we have to divide by $n!$.

We repeat that while for some x-s, π-s and l-s the equation (39) might have no solutions, for other data it can have more than one solution. Clearly, the set $\bar{X}\left(y \mid x_1, ..., x_{n-1}; \{l_1, ..., l_n\}\right)$, for Lebesgue-almost every data $x_1, ..., x_{n-1}$, can have no other entries than those of the form

$$x_{A,y,\{l_i\}} = y - \sum_{i \in A \subset \{1,2,...,n\}} l_i,$$

where A runs over all nonempty subsets of $\{1, 2, ..., n\}$ (i.e. at most $2^n - 1$ different entries). So the function $\bar{X}\left(y \mid x_1, ..., x_{n-1}; \{l_1, ..., l_n\}\right)$, as a function of $x_1, ..., x_{n-1}$, has to be piecewise constant. It is not ruled out apriori that for some data the set $\bar{X}\left(y \mid x_1, ..., x_{n-1}; \{l_1, ..., l_n\}\right)$ can be empty. This is not, however, the case. Moreover, as the crucial Theorem 4 below states,

- the number of elements in the set $\bar{X}\left(y \mid x_1, ..., x_{n-1}; \{l_1, ..., l_n\}\right)$, counted with multiplicities, is **precisely** $n!$ for Lebesgue-almost every value of the arguments.

Therefore we have for the inner integral in (38):

$$\int \frac{1}{n!} \lambda\left(\bar{X}\left(y \mid x_1, ..., x_{n-1}; \{l_1, ..., l_n\}\right)\right) \prod_{i=1}^{n-1} \left(\frac{\lambda(x_i)}{I_\lambda(y)} dx_i\right)$$
$$= \int \frac{1}{n!} \sum_{\substack{A \subset \{1,2,...,n\}, \\ A \neq \emptyset}} k\left(A, y, x_1, ..., x_{n-1}; \{l_1, ..., l_n\}\right) \lambda\left(x_{A,y,\{l_i\}}\right) \prod_{i=1}^{n-1} \left(\frac{\lambda(x_i)}{I_\lambda(y)} dx_i\right),$$

where the integer $k\left(A, y, x_1, ..., x_{n-1}; \{l_1, ..., l_n\}\right)$ is the multiplicity of the value $x_{A,y,\{l_i\}}$ of the function $\bar{X}$ at the point $\left(y, x_1, ..., x_{n-1}; \{l_1, ..., l_n\}\right)$. Rewriting it as

$$\int \frac{1}{n!} \lambda\left(\bar{X}\left(y \mid x_1, ..., x_{n-1}; \{l_1, ..., l_n\}\right)\right) \prod_{i=1}^{n-1} \left(\frac{\lambda(x_i)}{I_\lambda(y)} dx_i\right)$$
$$= \sum_{\substack{A \subset \{1,2,...,n\}, \\ A \neq \emptyset}} q_{\lambda,y}\left(A \mid \{l_1, ..., l_n\}\right) \lambda\left(x_{A,y,\{l_i\}}\right),$$

where

$$q_{\lambda,y}\left(A \,\middle|\, \{l_1,...,l_n\}\right) \tag{40}$$
$$= \int \frac{1}{n!} k\left(A,y,x_1,...,x_{n-1};\{l_1,...,l_n\}\right) \prod_{i=1}^{n-1}\left(\frac{\lambda(x_i)}{I_\lambda(y)}\,dx_i\right),$$

we have, due to the fact that Lebesgue-a.e.

$$\sum_{\substack{A\subset\{1,2,...,n\},\\ A\neq\emptyset}} k\left(A,y,x_1,...,x_{n-1};\{l_1,...,l_n\}\right) = n!,$$

the relations

$$0 \le q_{\lambda,y}\left(A \,\middle|\, \{l_1,...,l_n\}\right) \le 1, \text{ with } \sum_{\substack{A\subset\{1,2,...,n\},\\ A\neq\emptyset}} q_{\lambda,y}\left(A \,\middle|\, \{l_1,...,l_n\}\right) = 1, \tag{41}$$

since the measures $\frac{\lambda(x_i)}{I_\lambda(y)}\,dx_i$ are probability measures on $[0,y]$. (Note that the functions $k\left(A,y,x_1,...,x_{n-1};\{l_1,...,l_n\}\right)$ do depend on the variables $x_1,...,x_{n-1}$; hence the measures $q_{\lambda,y}\left(\cdot \,\middle|\, \{l_1,...,l_n\}\right)$ indeed depend on λ, y.) Therefore, for the function $b_n(y)$ we obtain a sort of a convolution expression:

$$b_n(y) = \int \sum_{\substack{A\subset\{1,2,...,n\},\\ A\neq\emptyset}} \left[q_{\lambda,y}\left(A \,\middle|\, \{l_1,...,l_n\}\right) \lambda\left(x_{A,y,\{l_i\}}\right)\right] \prod_{i=1}^{n} p(l_i)\;dl_i. \tag{42}$$

Be it the case that the probability measure $q_{\lambda,y}\left(\cdot \,\middle|\, \{l_1,...,l_n\}\right)$ is concentrated on just one subset $A = \{1,2,...,n\}$, we would obtain the usual convolution

$$b_n(y) = \int \lambda\left(y - l_1 - ... - l_n\right) \prod_{i=1}^{n} p(l_i)\;dl_i = \lambda * \underbrace{p * ... * p}_{n}(y).$$

Here the situation is more subtle, and in (42) we have a stochastic mixture of convolutions with random number of summands.

Taking into account the relations (34), (37), (42), the result can be summarized as follows. Let $\theta \equiv \theta_{\lambda,y}$ be the integer valued random variable with the distribution

$$\Pr\{\theta = n\} = \exp\{-I_\lambda(y)\}\frac{[I_\lambda(y)]^n}{n!},\; n = 0,1,2,...,$$

and $\eta_1, \eta_2, ...$ be the i.i.d. random serving times. Consider the random function $\xi_{\lambda,y} = \xi_{\lambda,y}(\theta_{\lambda,y}; \eta_1, \eta_2, ...)$, such that its conditional distribution under condition that the realization $\theta_{\lambda,y}; \eta_1, \eta_2, ...$ is given, is supported by the finite set

$$L(\theta_{\lambda,y}; \eta_1, \eta_2, ...) = \left\{ \sum_{i \in A} \eta_i : A \subset \{1, 2, ..., \theta_{\lambda,y}+1\}, A \neq \emptyset \right\} \subset \mathbb{R}^1,$$

and is given by

$$\Pr\left\{ \xi_{\lambda,y} = \sum_{i \in A} \eta_i \;\middle|\; \theta_{\lambda,y}; \eta_1, \eta_2, ... \right\} = q_{\lambda,y}\left(A \;\middle|\; \{\eta_1, ..., \eta_{\theta_{\lambda,y}+1}\} \right)$$

(see (40)). Then the following holds:

$$b(y) = \mathbb{E}(\lambda(y - \xi_{\lambda,y})).$$

This is precisely the relation (25), with $q_{\lambda,y}$ being the distribution of $\xi_{\lambda,y}$. The relation (30) follows directly from (41). ■

6 Combinatorics of the rod placements

In this section we will prove the Theorem 4, which was used in the previous section. We will use the notation of the previous section, introduced in the proof of the Theorem 3, up to relation (33).

By a cluster of the r-configuration $\sigma_n(z_1, ..., z_n; l_1, ..., l_n)$ with $z_1 < ... < z_n$ we call any maximal subsequence $z_i < z_{i+1} < ... < z_j$ such that $z_j = z_i + l_i + l_{i+1} + ... + l_{j-1}$. (The segment $[z_i, z_i + l_i + l_{i+1} + ... + l_j] \equiv [x_i, z_j + l_j]$ is what is called "busy period" for the queue.) If $z_i < z_{i+1} < ... < z_j$ is a cluster of an r-configuration, then the point z_i will be called the root of the cluster, while the point z_j will be called the head of the cluster. Note that for Lebesgue almost every configuration $\sigma_n(x_1, ..., x_n; l_1, ..., l_n)$ the point z_i is a root of a cluster of the corresponding r-configuration if and only if $z_i = x_i$. The segment $[z_i, z_j + l_j]$ will be called the body of the cluster $z_i < z_{i+1} < ... < z_j$, and the point $z_j + l_j$ will be called the end of the cluster.

The notation $\sigma_n(x_1, ..., x_n; l_1, ..., l_n) \cup \sigma_1(X, L)$ has the obvious meaning of adding an extra rod of the length L at the location X. Note though, that in general

$$R[\sigma_n(x_1, ..., x_n; l_1, ..., l_n) \cup \sigma_1(X, L)] \neq R[R\sigma_n(x_1, ..., x_n; l_1, ..., l_n) \cup \sigma_1(X, L)].$$

It is however the case, if the point X is outside the union of all bodies of clusters of $R\sigma_n(x_1, ..., x_n; l_1, ..., l_n)$. This will be used later.

In what follows we will need a marked point in $\mathbb{R}^1$. For all our purposes it is convenient to chose the origin, $0 \in \mathbb{R}^1$, as such a point.

We will say that the resolution of conflicts in the configuration $\sigma_n(x_1, ..., x_n; l_1, ..., l_n)$ results in a **hit** of the origin, iff for some k we have

$$y_k \equiv z_k + l_k = 0. \tag{43}$$

Such a hit will be called an x_r-hit, iff the cluster of the point z_k has its root at $z_r = x_r$. (Necessarily, we have that $r \leq k$.) An x_r-hit will be called an (x_r, x_k)-hit, if (43) holds.

Now we are ready to formulate our problem. Let n be an integer, and $\lambda_1 < \lambda_2 < ... < \lambda_n$ be a fixed set of positive lengths of rods. Let $x_1 < x_2 < ... < x_{n-1}$ be a set of $(n-1)$ left-ends. We want to compute the number $N(x_1, x_2, ..., x_{n-1}; \lambda_1, \lambda_2, ..., \lambda_n)$, which is defined as follows. For any permutation π of n elements and for any $X \in \mathbb{R}^1$, $X \neq x_1, x_2, ..., x_{n-1}$ we can consider the configuration $\sigma_{n-1}\left(x_1, ..., x_{n-1}; \lambda_{\pi(1)}, ..., \lambda_{\pi(n-1)}\right) \cup \sigma_1\left(X, \lambda_{\pi(n)}\right)$ of rods, when the rods $l_i = \lambda_{\pi(i)}$ are placed at $x_i, i = 1, ..., n-1$, while the free rod $l_n = \lambda_{\pi(n)}$ is placed at X. Given π, we count the number $N_\pi(x_1, ..., x_{n-1}; \lambda_1, ..., \lambda_n)$ of different locations X, such that the corresponding r-configuration $R\left[\sigma_{n-1}\left(x_1, ..., x_{n-1}; \lambda_{\pi(1)}, ..., \lambda_{\pi(n-1)}\right) \cup \sigma_1\left(X, \lambda_{\pi(n)}\right)\right]$ has a hit, and moreover this hit is an X-hit. (In certain cases one cannot produce an X-hit by putting the rod $l_n = \lambda_{\pi(n)}$ anywhere on $\mathbb{R}^1$; then $N_\pi(x_1, ..., x_{n-1}; \lambda_1, ..., \lambda_n) = 0$. In certain other cases there are more than one possibility to place the free rod so as to produce an X-hit.) Then we define

$$N(x_1, ..., x_{n-1}; \lambda_1, ..., \lambda_n) = \sum_{\pi \in \mathcal{S}_n} N_\pi(x_1, ..., x_{n-1}; \lambda_1, ..., \lambda_n).$$

Theorem 4 *For Lebesgue-almost every $x_1, ..., x_{n-1}$ and $\lambda_1, ..., \lambda_n$,*

$$N(x_1, ..., x_{n-1}; \lambda_1, ..., \lambda_n) = n!$$

Proof. Let us explain why the result is plausible. Let the set $x_1, ..., x_{n-1}$ be given. Then we can choose the positive numbers $\lambda_1, ..., \lambda_n$ so small that for any π the configuration

$\sigma_{n-1}\left(x_1,...,x_{n-1};\lambda_{\pi(1)},...,\lambda_{\pi(n-1)}\right)\cup\sigma_1\left(X=-\lambda_{\pi(n)},\lambda_{\pi(n)}\right)$, having the (X,X)-hit, has no conflicts, while no other choice of X results in a hit. Therefore in our case $N_\pi\left(x_1,...,x_{n-1};\lambda_1,...,\lambda_n\right)=1$ for every π, so indeed $N\left(x_1,...,x_{n-1};\lambda_1,...,\lambda_n\right)=n!$.

Now we explain why our result is non-trivial. To see it, take $n=2$, $x_1=-3$, $\lambda_1=1$, $\lambda_2=10$. Then

$$N_{12}\left(x_1;\lambda_1,\lambda_2\right)=2$$

– one can place the rod 10 at -10 or at -11. On the other hand,

$$N_{21}\left(x_1;\lambda_1,\lambda_2\right)=0$$

– the rod 10, placed at -3, blocks the origin from being hit. Still, $2+0=2!$. Note that this example is a general position one.

We will derive our theorem from its special case, explained in the first paragraph of the present proof. The idea of computing $N\left(x_1,...,x_{n-1};\lambda_1,...,\lambda_n\right)$ for a general data is to decrease one by one the numbers $\lambda_1<\lambda_2<...<\lambda_n$, starting from the smallest one, to the values very small, keeping track on the quantities $N_\pi\left(x_1,...,x_{n-1};\lambda_1,...,\lambda_n\right)$. During this evolution some of these will jump, but the total sum $N\left(x_1,...,x_{n-1};\lambda_1,...,\lambda_n\right)$ would stay unchanged, as we will show. That will prove our theorem.

We begin by presenting a simple formula for the number $N_\pi\left(x_1,...,x_{n-1};\lambda_1,..,\lambda_n\right)$. Consider the rod configuration $R\left[\sigma_{n-1}\left(x_1,...,x_{n-1};\lambda_{\pi(1)},...,\lambda_{\pi(n-1)}\right)\right]$, which will be abbreviated as $R_\pi\left(\lambda_1,..,\lambda_n\right)\equiv R_\pi\left(\lambda\right)$. Let us compute the quantity $S_\pi\left(x_1,...,x_{n-1};\lambda_1,..,\lambda_n\right)$, which is the number of points $y_i\in Y\left(R_\pi\left(\lambda_1,..,\lambda_n\right)\right)$, falling into the segment $\left[-\lambda_{\pi(n)},0\right]$.

Lemma 5

$$N_\pi\left(x_1,...,x_{n-1};\lambda_1,..,\lambda_n\right)=\begin{cases} S_\pi\left(x_1,...,x_{n-1};\lambda_1,..,\lambda_n\right) & \begin{array}{l}\textit{if the point } -\lambda_{\pi(n)}\\ \textit{belongs to a cluster}\\ \textit{of } R_\pi\left(\lambda_1,..,\lambda_n\right),\end{array} \\ S_\pi\left(x_1,...,x_{n-1};\lambda_1,..,\lambda_n\right)+1 & \textit{otherwice.}\end{cases} \tag{44}$$

Proof of the Lemma 5. Indeed, for every y_i, falling inside $\left[-\lambda_{\pi(n)},0\right]$, there is a position

$X_i(z_1, ..., z_{n-1}, y_1, ..., y_{n-1}) < 0$, such that once the free rod $\lambda_{\pi(n)}$ is placed there, the site y_i is pushed to the right and hits the origin. In case the point $-\lambda_{\pi(n)}$ is outside all clusters of $R_\pi(\lambda_1, .., \lambda_n)$, placing the free rod $\lambda_{\pi(n)}$ at $X_0 = -\lambda_{\pi(n)}$ produces an extra hit. ■

Now let $\Delta > 0$ be such that

$$\lambda_1 < \lambda_2 < ... < \lambda_{i-1} < \lambda_i - \Delta < \lambda_i + \Delta < \lambda_{i+1} < ... < \lambda_n$$

for some $i = 1, ..., n$, and some of the functions N_π exhibit jumps in the variable λ_i as it goes down from $\lambda_i + \Delta$ to $\lambda_i - \Delta$. We denote by $\lambda(\delta)$ the vector $\lambda_1, ..., \lambda_i + \delta, ..., \lambda_n$. We suppose that Δ is small enough, so that for any π the difference

$$|N_\pi(x_1, ..., x_{n-1}; \lambda(\Delta)) - N_\pi(x_1, ..., x_{n-1}; \lambda(-\Delta))|$$

is at most one. Moreover, we want Δ to be so small that on the segment $\lambda \in [\lambda_i - \Delta, \lambda_i + \Delta]$ there is precisely one point, say λ_i, at which some of the functions $N_\pi(x_1, ..., x_{n-1}; \lambda)$ do jump. (In general, there will be several permutations π, for which such a jump will happen at $\lambda = \lambda_i$. Indeed, if we observe an (X, x_k)-hit in our rod configuration with $l_i = \lambda_{\pi(i)}$, while we have that $x_1 < x_2 < ...x_{s-1} < X < x_s < ... < x_k < ... < x_{n-1}$, then in some cases we will have an (X, x_k)-hit for every rearrangement of the rods $l_s, ..., l_k$, i.e. for all permutations of the form $\pi \circ \rho$, where ρ permutes the elements $s, ..., k$, leaving the other fixed, see Lemma 5.)

Let us begin with the case when

$$N_\pi(x_1, ..., x_{n-1}; \lambda(\Delta)) - N_\pi(x_1, ..., x_{n-1}; \lambda(-\Delta)) = 1. \tag{45}$$

That means that either the intersection $Y(R_\pi(\lambda(\Delta))) \cap \left[-\lambda(\Delta)_{\pi(n)}, 0\right]$ is non-empty and after the 2δ-evolution its cardinality decreases by one, or else that the point $-\lambda(\Delta)_{\pi(n)}$ is outside all clusters of $R_\pi(\lambda(\Delta))$, while the point $-\lambda(-\Delta)_{\pi(n)}$ is inside some cluster of $R_\pi(\lambda(-\Delta))$. In the first case let $y_k(\lambda(\Delta), \pi) < ... < y_r(\lambda(\Delta), \pi)$ be all the points of the above intersection. The relation (45) implies via (44) that the point $y_k(\lambda(\delta), \pi)$ leaves the segment $\left[-\lambda(\delta)_{\pi(n)}, 0\right]$ as δ passes the zero value:

$$y_k(\lambda(\delta), \pi) > -\lambda(\delta)_{\pi(n)} \text{ for } \delta > 0, \tag{46}$$

$$y_k(\lambda(0), \pi) = -\lambda(0)_{\pi(n)}, \tag{47}$$

$$y_k(\lambda(\delta), \pi) < -\lambda(\delta)_{\pi(n)} \text{ for } \delta < 0. \tag{48}$$

Moreover, the point $y_k(\lambda(\delta), \pi)$ is not the end of the cluster – otherwise we would have $N_\pi(x_1, ..., x_{n-1}; \lambda(\Delta)) = N_\pi(x_1, ..., x_{n-1}; \lambda(-\Delta))$. Therefore $y_k(\lambda(\delta), \pi) = z_{k+1}(\lambda(\delta), \pi)$. We now claim that if we assign the rod $\lambda(\delta)_{\pi(n)}$ to x_{k+1}, and will take for the free rod the rod $\lambda(\delta)_{\pi(k+1)}$, then for the corresponding permutation the opposite to (45) happens:

$$N_{\pi\circ(n\leftrightarrow k+1)}(x_1, ..., x_{n-1}; \lambda(\Delta)) - N_{\pi\circ(n\leftrightarrow k+1)}(x_1, ..., x_{n-1}; \lambda(-\Delta)) = -1. \tag{49}$$

(Here we denote by $\pi \circ (n \leftrightarrow k+1)$ the permutation which is the composition of the transposition $n \leftrightarrow k+1$, followed by π.) Indeed, after the above reassignment and the resolution of conflicts, the rod $\lambda(\delta)_{\pi(n)}$ will be positioned at the point $y_k(\lambda(\delta), \pi)$. The relations $(46) - (48)$ then tell us that during the 2δ-evolution the right endpoint of this rod will move from the positive semiaxis to the negative one, thus adding one unit to the value $S_{\pi\circ(n\leftrightarrow k+1)}(x_1, ..., x_{n-1}; \lambda(\Delta))$.

In the second case we have that the point $-\lambda(\Delta)_{\pi(n)}$ is outside all clusters of $R_\pi(\lambda(\Delta))$, while the point $-\lambda(-\Delta)_{\pi(n)}$ is inside a cluster of $R_\pi(\lambda(-\Delta))$. That however can happen only if $\pi(n) = i$, so the rod $\lambda(\delta)_{\pi(n)}$ itself is varying with δ, and in addition to it we have for some $k \neq n$ that

$$-\lambda(\delta)_{\pi(n)} \begin{cases} < x_k & \text{if } \delta > 0, \\ = x_k & \text{if } \delta = 0, \\ > x_k & \text{otherwice,} \end{cases}$$

while $\lambda_{\pi(k)} > -x_k$. But then for a permutation $\pi \circ (n \leftrightarrow k)$ we have that the variable rod $\lambda_i(\delta)$ is assigned to the location x_k, is not relocated due to the resolution of conflicts, so the endpoint y_k equals $x_k + \lambda_i(\delta)$, and during the evolution crosses the zero threshold from right to left; therefore again

$$N_{\pi\circ(n\leftrightarrow k)}(x_1, ..., x_{n-1}; \lambda(\Delta)) - N_{\pi\circ(n\leftrightarrow k)}(x_1, ..., x_{n-1}; \lambda(-\Delta)) = -1.$$

The above construction corresponds to every permutation π, satisfying (45), another permutation, $\pi' = \Phi(\pi)$, which satisfy (49). We will be done if we show that Φ is one to one. We prove this by constructing the inverse of Φ.

So let π' be such that

$$N_{\pi'}(x_1, ..., x_{n-1}; \lambda(\Delta)) - N_{\pi'}(x_1, ..., x_{n-1}; \lambda(-\Delta)) = -1.$$

That means that the cardinality of the intersection $Y\left(R_{\pi'}\left(\lambda\left(\Delta\right)\right)\right)\cap\left[-\lambda\left(\Delta\right)_{\pi'(n)},0\right]$ increases by one during the 2δ-evolution: indeed, the only other option, due to the Lemma 5, is that the point $-\lambda\left(\Delta\right)_{\pi'(n)}$ is inside some cluster of $R_{\pi'}\left(\lambda\left(\Delta\right)\right)$, while the point $-\lambda\left(-\Delta\right)_{\pi'(n)}$ is outside all clusters of $R_{\pi'}\left(\lambda\left(-\Delta\right)\right)$. But in this case $N_{\pi'}\left(x_1,...,x_{n-1};\lambda\left(\Delta\right)\right)=N_{\pi'}\left(x_1,...,x_{n-1};\lambda\left(-\Delta\right)\right)$.

In our situation we necessarily have that the intersection $Y\left(R_{\pi'}\left(\lambda\left(\Delta\right)\right)\right)\cap\left(0,+\infty\right)\neq\emptyset$. Let $y_{k'}\left(\lambda\left(\Delta\right),\pi'\right)<...<y_{r'}\left(\lambda\left(\Delta\right),\pi'\right)$ are all the points of this intersection. The relation (45) implies via (44) that the point $y_{k'}\left(\lambda\left(\delta\right),\pi'\right)$ moves from the positive semiaxis to the negative one as δ passes the zero value:

$$y_{k'}\left(\lambda\left(\delta\right),\pi'\right)>0 \text{ for } \delta>0, \tag{50}$$

$$y_{k'}\left(\lambda\left(0\right),\pi'\right)=0, \tag{51}$$

$$y_{k'}\left(\lambda\left(\delta\right),\pi'\right)<0 \text{ for } \delta<0. \tag{52}$$

Two subcases are possible. The first one is when the point $z_{k'}$ is the head of the cluster, so $z_{k'}=x_{k'}$, and the rod $\lambda_{\pi'(k')}$ is the variable one. Then for the permutation $\pi''=\pi'\circ\left(n\leftrightarrow k'\right)$ the following happen: the free rod $\lambda_{\pi''(n)}$ is the variable one, and as δ varies, the point $-\lambda_{\pi''(n)}$ moves from the location to the left of $x_{k'}$ to the location to the right to $x_{k'}$. Since the point $x_{k'}$ is the head of the cluster, we have therefore that

$$N_{\pi''}\left(x_1,...,x_{n-1};\lambda\left(\Delta\right)\right)-N_{\pi''}\left(x_1,...,x_{n-1};\lambda\left(-\Delta\right)\right)=1. \tag{53}$$

In the second subcase $z_{k'}=y_{k'-1}$. Then the relations $(50)-(52)$ mean that the point $y_{k'-1}\left(\lambda\left(\delta\right),\pi'\right)$ is inside the segment $\left[-\lambda\left(\delta\right)_{\pi'(k')},0\right]$ for $\delta=\Delta$, and outside it for $\delta=-\Delta$. So if we again assign the free rod $\lambda\left(\delta\right)_{\pi'(n)}$ to the point $x_{k'}$, making instead the rod $\lambda\left(\delta\right)_{\pi'(k')}$ to be free, then for the permutation $\pi''=\pi'\circ\left(n\leftrightarrow k'\right)=\Phi'\left(\pi'\right)$ the relation (53) again holds.

The statement that Φ' is inverse to Φ is straightforward. ∎

Below we will need a version of the above theorem, which follows. Let T,L be positive real, $L<T$. Let again n be an integer, and $\lambda_1<\lambda_2<...<\lambda_n$ be a fixed set of positive lengths of rods. Let $-T<x_1<x_2<...<x_{n-1}<0$ be a set of $(n-1)$ left-ends. We want to compute the number $\tilde{N}\left(-T,x_1,x_2,...,x_{n-1};L,\lambda_1,\lambda_2,...,\lambda_n\right)$, which is defined as follows. For any permutation π of n elements and for any $X\in\left(-T,0\right)$, $X\neq x_1,x_2,...,x_{n-1}$ we

can consider the configuration $\sigma_n\left(-T, x_1, ..., x_{n-1}; L, \lambda_{\pi(1)}, ..., \lambda_{\pi(n-1)}\right) \cup \sigma_1\left(X, \lambda_{\pi(n)}\right)$ of rods, when the rod L is placed at $-T$, the rods $l_i = \lambda_{\pi(i)}$ are placed at $x_i, i = 1, ..., n-1$, while the free rod $l_n = \lambda_{\pi(n)}$ is placed at X, $-T < X < 0$. Given π, we count the number $\tilde{N}_\pi\left(-T, x_1, x_2, ..., x_{n-1}; L, \lambda_1, \lambda_2, ..., \lambda_n\right)$ of different locations X, such that the corresponding r-configuration $R\left[\sigma_n\left(-T, x_1, ..., x_{n-1}; L, \lambda_{\pi(1)}, ..., \lambda_{\pi(n-1)}\right) \cup \sigma_1\left(X, \lambda_{\pi(n)}\right)\right]$ has a hit at zero, and moreover this hit is an X-hit. Then we define

$$\tilde{N}\left(-T, x_1, x_2, ..., x_{n-1}; L, \lambda_1, \lambda_2, ..., \lambda_n\right) = \sum_{\pi \in \mathcal{S}_n} \tilde{N}_\pi\left(-T, x_1, x_2, ..., x_{n-1}; L, \lambda_1, \lambda_2, ..., \lambda_n\right).$$

Theorem 6 *Suppose that*

$$L + \lambda_1 + \lambda_2 + ... + \lambda_n < T. \tag{54}$$

Then $\tilde{N}\left(-T, x_1, x_2, ..., x_{n-1}; L, \lambda_1, \lambda_2, ..., \lambda_n\right) = n!$ *for Lebesgue-almost every* $x_1, ..., x_{n-1}$ *and* $\lambda_1, ..., \lambda_n$.

The Theorem 6 differs from the Theorem 4 by the presence of the additional rod L, which is placed at $-T$, and by the restriction that all points $X, x_1, x_2, ..., x_{n-1}$ has to be within the segment $(-T, 0)$. In particular, the rod L does not have to move under the resolution of conflicts.

Note that without the restriction (54) the statement of the theorem is not valid, as it is easy to see.

Proof. Let the numbers $0 < \varepsilon_1 < ... < \varepsilon_{n-1}$ be so small that the sum $\varepsilon_1 + ... + \varepsilon_{n-1}$ is less than any of the numbers $\left|\delta_0(T-L) + \delta_1\lambda_1 + \delta_2\lambda_2 + ... + \delta_n\lambda_n\right|$, where δ_i are taking any of three values $-1, 0, 1$, with the only restriction that not all of them vanish simultaneously. Let us replace the configuration $x_1, x_2, ..., x_{n-1}$ by the configuration $x_1', x_2', ..., x_{n-1}'$, where

$$x_i' = \begin{cases} L - T + \varepsilon_i & \text{if } x_i < L - T, \\ x_i & \text{otherwice.} \end{cases}$$

Let k be the largest integer for which $x_i' > x_i$. (The meaning of the configuration $x_1', x_2', ..., x_{n-1}'$ is the following: were all ε_i zeroes, it is the result of resolving the first conflict, between the first rod L and the rods intersecting it, which rods have to be pushed to the right-hand end of L. We use positive ε-s in order to have all the point x_i' different.) By the previous theorem we know that $N\left(x_1', x_2', ..., x_{n-1}'; \lambda_1, ..., \lambda_n\right) = n!$. Let the location X and the permutation π are such that the corresponding r-configuration

$R\left[\sigma_{n-1}\left(x'_1,...,x'_{n-1};\lambda_{\pi(1)},...,\lambda_{\pi(n-1)}\right)\cup\sigma_1\left(X,\lambda_{\pi(n)}\right)\right]$ has an X-hit. The condition (54) implies that the cluster of the r-configuration $R\left[\sigma_{n-1}\left(x'_1,...,x'_{n-1};\lambda_{\pi(1)},...,\lambda_{\pi(n-1)}\right)\cup\sigma_1\left(X,\lambda_{\pi(n)}\right)\right]$, rooted at X, does not contain any of the points $z'_1=x'_1,z'_2,...,z'_k$ (see (31) for the notation), so $X>L-T$, and the r-configuration $R\left[\sigma_n\left(-T,x_1,...,x_{n-1};L,\lambda_{\pi(1)},...,\lambda_{\pi(n-1)}\right)\cup\sigma_1\left(X,\lambda_{\pi(n)}\right)\right]$ has an X-hit as well. Therefore $\tilde{N}\left(-T,x_1,x_2,...,x_{n-1};L,\lambda_1,\lambda_2,...,\lambda_n\right)\geq n!$. On the other hand, if the r-configuration $R\left[\sigma_n\left(-T,x_1,...,x_{n-1};L,\lambda_{\pi(1)},...,\lambda_{\pi(n-1)}\right)\cup\sigma_1\left(X,\lambda_{\pi(n)}\right)\right]$ has an X-hit, then by the same reasoning X has to be to the right of the location $L-T$, and moreover the cluster of this configuration, rooted at X, does not contain any of the points $z_1=-T,z_2=-T+L,...,z_{k+1}$; therefore the r-configuration $R\left[\sigma_{n-1}\left(x'_1,...,x'_{n-1};\lambda_{\pi(1)},...,\lambda_{\pi(n-1)}\right)\cup\sigma_1\left(X,\lambda_{\pi(n)}\right)\right]$ has an X-hit. Hence $\tilde{N}\left(-T,x_1,x_2,...,x_{n-1};L,\lambda_1,\lambda_2,...,\lambda_n\right)\leq n!$, and the proof follows. ■

7 10 (ten) technical statements

In this section we present several technical statements needed for the proof of our main result. The first subsection deals with the regularity properties of the NMP-s, while the second – with the estimates on the averaging kernels $q_{*,*}$.

7.1 Regularity properties of Non-Linear Markov Process

The first fact we will establish concerns the integral behavior of the output rate (=input rate) of the NMP.

Lemma 7 *Let $\mu_{\nu,\lambda_\nu(\cdot)}(\cdot)$ be NMP, with $N\left(\mu_{\nu,\lambda_\nu(\cdot)}(t)\right)=N(\nu)=q$. Then there exists a time duration $T'=T'(q)$ and $\varepsilon'=\varepsilon'(q)>0$, such that for all $T\geq T'$ and all $s\geq 0$*

$$\int_s^{s+T}\lambda_\nu(t)\,dt<T\left(1-\varepsilon'\right). \tag{55}$$

Clearly, in order to see it, it is sufficient to prove the following statement:

Proposition 8 *Let μ_t be GFP (i.e. time inhomogeneous Markov process M(t)/GI/1, see Sect. 2) with arbitrary initial state μ_0, corresponding to the Poisson input with continuous rate $\lambda(t)$, $0 \leq \lambda(t) \leq L$. For any $N > 0$ one can find $\varepsilon(N)$ and $T(N, L)$, such that if for some $T \geq T(N, L)$*

$$\int_0^T \lambda(t)\, dt \geq (1 - \varepsilon(N))\, T, \tag{56}$$

then for some $t \leq T$

$$N(\mu_t) \geq N. \tag{57}$$

Indeed, if for any T and ε' one can find s such that (55) is violated, then Proposition 8, used with $L = C(\eta)$ (see (17)), would imply that the function $N\left(\mu_{\nu,\lambda_\nu(\cdot)}(t)\right)$ is unbounded. This, however, contradicts to the fact that it stays constant. The continuity of $\lambda_\nu(t)$, which is a prerequisite needed to apply Proposition 8, is a consequence of Lemma 14 below.

The proof of the Proposition 8 (which proposition is of course valid even without continuity assumption) follows from the next two lemmas, preceded by two definitions.

Let χ_1, χ_2 be two measures on a segment $[A, B] \subset \mathbb{R}^1$.

Definition 9 *We say that $\chi_1 \prec \chi_2$, if for any monotone increasing function f on $[A, B]$ we have*

$$\int_A^B f\ d\chi_1 \leq \int_A^B f\ d\chi_2.$$

(This is equivalent to saying that $\chi_1([a, B]) \leq \chi_2([a, B])$ for every $b \in [A, B]$.)

Definition 10 *We say that $\chi_1 \leq \chi_2$ if for every positive function g on $[A, B]$ we have*

$$\int_A^B g\ d\chi_1 \leq \int_A^B g\ d\chi_2.$$

(This is equivalent to saying that $\chi_1([a, b]) \leq \chi_2([a, b])$ for every $a, b \in [A, B]$.) Note that the second relation holds for probability measures only in case when $\chi_1 = \chi_2$; in this paragraph we are concerned, however, with arbitrary measures.

In what follows we will assume that the measures χ_i have densities λ_i with respect to the Lebesgue measure, which densities are continuous and satisfy $0 \leq \lambda_i \leq C$, though much of what will be said is true in more general situation.

Lemma 11 *Let us consider two GFP (non-stationary Markov processes) μ_t^1, μ_t^2, with the same distribution of the service time η, and with the Poisson input flows, defined by the two measures χ_1, χ_2 on the time interval $[A, B]$. Suppose that $\mu_A^1(\mathbf{0}) = 1$ (that is, the first server is initially idle), and that $\chi_1 \prec \chi_2$. Then*

$$N\left(\mu_B^1\right) \leq N\left(\mu_B^2\right).$$

Proof. The rough idea of the proof is the following: we will argue that the condition $\chi_1 \prec \chi_2$ enables us to represent the process μ_t^2 as a certain transformation of the process μ_t^1, when the customers of the second process are the same as these in the first process (i.e. they require the same service times), but just come later, in addition to extra customers which were not present in the first process. Clearly, in that case the queue at the final moment has to be longer for the second process.

To make the above rigorous, we will use the coupling technique. Note first, that if $\mathcal{P}$ is a Poisson random field on $[A, B]$ with the rate function $\lambda(t)$, $t \in [A, B]$, $\omega \subset [A, B]$ is its realization, and $f : [A, B] \to [A, B]$ is a strictly increasing continuous map, then the image set $f(\omega)$ is also a realization of a Poisson random field $\mathcal{P}^f$, defined by the rate function $\lambda^f(t)$, given by

$$\lambda^f(t) = \begin{cases} \frac{\lambda\left(f^{-1}(t)\right)}{f'\left(f^{-1}(t)\right)} & \text{if } t \in f([A, B]), \\ 0 & \text{otherwice.} \end{cases}$$

We claim now that if $\chi_1 \prec \chi_2$, then there exists a map $f : [A, B] \to [A, B]$, such that

$$f(x) \geq x \text{ for all } x \in [A, B],$$
$$\lambda_1^f(t) \leq \lambda_2(t).$$

One way to construct such a function is the following. Let us extend the function $\lambda_2(t)$ to the region $t \geq B$ by putting it there to be equal to $\max \lambda_1$. Consider now the family Φ of all continuous functions $\varphi : [A, B] \to [A, \infty)$, satisfying the properties:

$$\varphi(x) \geq x;$$
$$\varphi(y) \geq \varphi(x) \text{ for } y > x;$$
$$\frac{\lambda_1(x)}{\varphi'(x)} \leq \lambda_2(\varphi(x)).$$

For example, any shift $\varphi_c(x) = x + c$ is in Φ, once $c > B - A$, so Φ is nonempty. Note that if $\varphi_1, \varphi_2 \in \Phi$, then the function $\tilde{\varphi}(x) = \min\{\varphi_1(x), \varphi_2(x)\}$ is also in Φ. Therefore the function

$$f(x) = \inf_{\varphi \in \Phi} \{\varphi(x)\}$$

is in Φ as well, and it is easy to see that $f([A, B]) \subset [A, B]$. (In fact, if $\chi_2([A, B]) = \chi_1([A, B])$, then f is the only function in Φ with this property.)

Another, more intuitive way of defining f, is by constructing a coupling between the measure χ_1 and a suitable measure $\chi_3 \succ \chi_1$, which is a "part" of χ_2, in the sense that $\chi_3 \leq \chi_2$. (The measure K on $[A, B] \times [A, B]$ is called a coupling between the measures χ' and χ'' on $[A, B]$ iff for every subset $C \subset [A, B]$ we have $K(C \times [A, B]) = \chi'(C)$ and $K([A, B] \times C) = \chi''(C)$.)

The special coupling we need can be most easily constructed via discrete approximations of the measures χ_i and subsequent limit procedure. Since this construction is well-known due to extensive use in probability theory of the Monge-Kantorovich-Rubinstein-Ornstein-Vasserstein distance, we will give only a sketch of it. We replace the segment $[A, B]$ by the set $\{1, 2, ..., n\} \subset \mathbb{R}^1$. The coupling sought is then just a matrix $K(i, j)$. We construct it by induction.

We define the first row $K(1, \cdot) = \{K(1, 1), K(1, 2), ..., K(1, n)\}$ to be the measure on $\{1, 2, ..., n\}$ with the following properties:

i) $K(1, 1) + K(1, 2) + ... + K(1, n) = \chi_1(1)$;

ii) $K(1, \cdot) \leq \chi_2(\cdot)$;

iii) $K(1, \cdot)$ is the minimal (in the $\prec$-sense) measure on $\{1, 2, ..., n\}$, satisfying i) and ii).

(One can give more explicit definition of K: we start by putting $K(1, 1) = \min\{\chi_1(1), \chi_2(1)\}$. If it turns out that $\chi_1(1) = K(1, 1)$, then we put $K(1, j) = 0$ for $j > 1$. Otherwise we define $K(1, 2) = \min\{\chi_1(1) - K(1, 1), \chi_2(2)\}$. If it turns out that $\chi_1(1) = K(1, 1) + K(1, 2)$, then we put $K(1, j) = 0$ for $j > 2$, otherwise defining $K(1, 3) = \min\{\chi_1(1) - K(1, 1) - K(1, 2), \chi_2(3)\}$, etc.)

Because of ii), the difference $\chi_2^{(2)}(\cdot) \equiv \chi_2(\cdot) - K(1, \cdot)$ is still a (positive) measure on $\{1, 2, ..., n\}$, and, because of iii), $\chi_2^{(2)}(\cdot) \succ \chi_1(\cdot)$, *considered as the measures on* $\{2, 3, ..., n\}$. Therefore we can repeat the preceding construction, defining the second row, $K(2, \cdot)$, as the measure on $\{2, 3, ..., n\}$, corresponding to $\chi_2^{(2)}(\cdot)$ and $\chi_1 \Big|_{\{2,3,\ldots,n\}}$. Proceeding inductively, and consid-

ering the measures $\chi_2^{(k)}(\cdot)$ on $\{k, k+1, ..., n\}$, we obtain the coupling $K(i,j)$ sought. By construction, $K(i,j) = 0$ for $j < i$. Finally, we define the measure χ_3 by

$$\chi_3(j) = K(1,j) + ... + K(j,j).$$

By construction, $\chi_3 \succ \chi_1$, while $\chi_3 \leq \chi_2$.

To perform the limiting procedure, we now define for every n the atomic measures $\chi_{i,n}$ with atoms at points $\left\{A + k\frac{B-A}{n},\ k = 0, 1, ..., n-1\right\}$, by

$$\chi_{i,n}\left(A + k\frac{B-A}{n}\right) = \chi_i\left(\left[A + k\frac{B-A}{n}, A + (k+1)\frac{B-A}{n}\right]\right).$$

We then construct the couplings K_n between $\chi_{1,n}$ and $\chi_{3,n}$ (with $\chi_{1,n} \prec \chi_{3,n},\ \chi_{3,n} \leq \chi_{2,n}$) in the manner described above. As $n \to \infty$, the supports $S_n = \operatorname{supp}(K_n) \subset [A,B] \times [A,B]$ of the measures K_n converge to the limiting curve S in the square $[A,B] \times [A,B]$. This curve is the graph of the above function f:

$$S = \{(x,y) : y = f(x)\}.$$

Now we can return to the proof of the statement of the Lemma. Let $\omega \in [A,B]$ be a configuration of the Poisson random field $\mathcal{P}_1$, corresponding to the rate measure χ_1. Let f be the function defined above, $\chi_3 = f_*(\chi_1)$, and the measure $\zeta = \chi_2 - \chi_3$ (that ζ is indeed a positive measure follows from our construction). Let $\tilde{\mathcal{P}}$ be the Poisson random field with the rate ζ, independent of $\mathcal{P}_1$, and $\tilde{\omega}$ be its configuration. Then the random set $\bar{\omega} = f(\omega) \cup \tilde{\omega}$ has distribution of the configuration of the Poisson random field $\mathcal{P}_2$, corresponding to the rate measure χ_2. To specify the input flow and thus the coupling sought, we have to specify the service times. So we assign to every point x in ω the service time η_x, drawn independently from the distribution of the random variable η. To every point $y \in f(\omega)$ we assign the service time $\eta_{f^{-1}(y)}$, while to points $z \in \tilde{\omega}$ we assign independent realizations η_z of η.

Our statement now becomes almost evident. Consider a $\mathcal{P}_1$-customer, represented by a point $x \in \omega$, whose service is not yet over at the moment B. This happens due to the service time η_x, needed for him, as well as due to the service of the customers $x_1, x_2, ...$, who came before x. But then the corresponding $\mathcal{P}_2$-customer $f(x)$ has the same service time η_x, but arrives later than x, as well as all the customers $f(x_1), f(x_2), ...$, who came before. In addition, there can be extra $\tilde{\mathcal{P}}$-customers arrived before $f(x)$. It is evident

therefore, that the service of the $\mathcal{P}_2$-customer $f(x)$ will not be over at the moment B. So the queue in the second case is not shorter. $\blacksquare$

In what follows we denote by $\kappa^{(a)}$ the measure on $[A,B]$, having the constant density $\lambda(t) = a$.

Lemma 12 *Let the measure χ on $[A,B]$ has continuous density $\lambda(t)$, $0 \leq \lambda(t) \leq L$, and satisfies the property:*

$$\chi([A,B]) \geq (1-\varepsilon)(B-A).$$

Then there exists a segment $[A,C] \subset [A,B]$ of the length

$$C - A > \frac{\varepsilon}{L}(B-A), \tag{58}$$

such that

$$\chi\Big|_{[A,C]} \succ \kappa^{(1-2\varepsilon)}\Big|_{[A,C]}.$$

Proof. Consider all the segments $[c,d] \subset [A,B]$, which have the property that

$$\chi\Big|_{[c,d]} \succ \kappa^{(1-2\varepsilon)}\Big|_{[c,d]}. \tag{59}$$

Note that if $[c_1,d_1]$ and $[c_2,d_2]$ are two such segments, and $[c_1,d_1] \cap [c_2,d_2] \neq \varnothing$, then the same holds for their union, i.e.

$$\chi\Big|_{[c_3,d_3]} \succ \kappa^{(1-2\varepsilon)}\Big|_{[c_3,d_3]}$$

for $[c_3,d_3] = [c_1,d_1] \cup [c_2,d_2]$. Therefore the union of all segments with the property (59) splits into the family of non-intersecting maximal ones, $[C_1,D_1], [C_2,D_2], \ldots$, where we enumerate the segments according to their length, say, so $|D_1 - C_1| \geq |D_2 - C_2| \geq \ldots$. Denote by $[A,a]$ the segment $[C_i,D_i]$ among these maximal, which contains the point A. If in fact A does not belong to any maximal segment, we put $a = A$.

Let now $[C_k,D_k]$ be one of maximal segments, which is different from the segment $[A,a]$. Then

$$\chi([C_k,D_k]) = (1-2\varepsilon)(D_k - C_k). \tag{60}$$

Indeed, let us couple the measure $\kappa^{(1-2\varepsilon)}\Big|_{[C_k,D_k]}$ with (part of) the measure $\chi\Big|_{[C_k,D_k]}$ – that is with some measure $\tilde{\chi}\Big|_{[C_k,D_k]} \leq \chi\Big|_{[C_k,D_k]}$, see the proof of

the Lemma 11 above. This is possible since $\kappa^{(1-2\varepsilon)}\Big|_{[C_k,D_k]} \prec \chi\Big|_{[C_k,D_k]}$. If $\chi([C_k,D_k]) > (1-2\varepsilon)(D_k - C_k)$, then the difference $\Delta\chi\Big|_{[C_k,D_k]} = \chi\Big|_{[C_k,D_k]} - \tilde{\chi}\Big|_{[C_k,D_k]}$ is a positive measure on $[C_k, D_k]$. Therefore for some small δ the measure $\kappa^{(1-2\varepsilon)}\Big|_{[C_k-\delta,C_k]}$ can be coupled with (part of) the measure $\Delta\chi\Big|_{[C_k,D_k]}$. That implies that $\chi\Big|_{[C_k-\delta,D_k]} \succ \kappa^{(1-2\varepsilon)}\Big|_{[C_k-\delta,D_k]}$, which contradicts to the maximality of $[C_k, D_k]$.

Note that for any point $x \in [A,B]$, which is outside all of the segments $[C_i, D_i]$, we have $\lambda(x) \leq 1 - 2\varepsilon$. Together with (60) it implies that

$$\chi([a,B]) \leq (1-2\varepsilon)|B-a|.$$

On the other hand,

$$\chi([A,B]) = \chi([A,a]) + \chi([a,B]) \geq (1-\varepsilon)|B-A|,$$

so

$$(a-A)L \geq \chi([A,a]) \geq \varepsilon|B-A|,$$

and the proof follows. ■

Proof of the Proposition 8. Let N be fixed. As we know (see relation (18) and the statement after it), there is a value $c = c(N) < 1$, such that the homogeneous process with the rate function $\lambda \equiv c$ has the invariant measure ν_c with $N(\nu_c) = 2N$. We define $\varepsilon(N)$ by

$$\varepsilon(N) = \frac{1-c(N)}{2}.$$

If we start the process with $\lambda \equiv c$ in the state $\mathbf{0}$, then it weakly converges with time to ν_c. In particular, $N(\mu_{\mathbf{0},c}(t)) \to 2N$, as $t \to \infty$. Define $\bar{T} = \bar{T}(N)$ to be the time duration, after which the (monotone) function $N(\mu_{\mathbf{0},c}(t))$ satisfies $N(\mu_{\mathbf{0},c}(t)) \geq N$ for all $t \geq \bar{T}$. We want this value $\bar{T}$ to appear as the lower bound of the length of the segment $[A.C]$, obtained in the Lemma 12. That will be the case if the whole segment $[A,B]$ will be of length

$$T(N,L) = \frac{L\bar{T}(N)}{\varepsilon(N)},$$

see (58). Now, if (56) is satisfied with the $T(N,L)$ and $\varepsilon(N)$ just chosen, then by Lemma 12 the measure $\lambda(t)\,dt$ is bigger (in the $\succ$ sense) than the measure $c(N)\,dt$ on a segment $[0,\bar{t}]$ with $\bar{t} \geq \bar{T}$. By Lemma 11 we have the relation (57) at $\bar{t}$. ■

Next we show that for every initial state ν of NMP there exists a time moment after which the probability of observing the system $\mu_{\nu,\lambda_\nu(\cdot)}(t)$ to be in the idle state is uniformly positive.

Lemma 13 *Let $\mu_{\nu,\lambda_\nu(\cdot)}(\cdot)$ be NMP, with $N\left(\mu_{\nu,\lambda_\nu(\cdot)}(t)\right) = N(\nu) = q$. Then there exists a time moment $T = T(\nu)$ and $\varepsilon = \varepsilon(\nu) > 0$, such that for all $t > T$*

$$\langle \omega = \mathbf{0} \rangle_{\mu_{\nu,\lambda_\nu(\cdot)}(t)} > \varepsilon. \tag{61}$$

Proof. We first construct an auxiliary stationary ergodic Markov process, which in a certain sense dominates our NMP from above. Namely, let $T = T'(q)$ be the time duration proven to exist in Lemma 7, and ε be the corresponding quantity $\varepsilon'(q)$. Our Markov process $M_{\nu,\varepsilon}(t)$ will consist of the states of some auxiliary server $\mathfrak{S}$ at moments t, with $M_{\nu,\varepsilon}(0) = \nu$. The customers are arriving to $\mathfrak{S}$ only at moments kT, $k = 1, 2, \ldots$. Their numbers $\mathfrak{N}_k$ are i.i.d., distributed as the number of Poisson flow of customers with constant rate $(1-\varepsilon)$, arriving during the time intervals $[(k-1)T, kT]$. The service times are described by our random variable η. Since $\mathbb{E}(\eta) = 1$, while $\varepsilon > 0$, $M_{\nu,\varepsilon}$ is indeed ergodic. In particular,

$$L(\varepsilon,\delta) \equiv \lim_{k\to\infty} \langle \omega = \mathbf{0} \rangle_{M_{\nu,\varepsilon}(kT-\delta)} > 0 \tag{62}$$

for any δ small enough.

Now we note that

$$\langle \omega = \mathbf{0} \rangle_{\mu_{\nu,\lambda_\nu(\cdot)}(kT-\delta)} \geq \Pr(\mathcal{E}_k(T)) \left\langle \omega = \mathbf{0} \,\middle|\, \mathcal{E}_k(T) \right\rangle_{\mu_{\nu,\lambda_\nu(\cdot)}(kT-\delta)}, \tag{63}$$

where the event

$$\mathcal{E}_k(T) = \left\{ \begin{array}{l} \text{in the Poisson random flow, defined by the rate} \\ \lambda_\nu(\cdot)\,,\ \text{no customer arrives during the time } [(k-1)T, kT]\,. \end{array} \right\}.$$

Since the rate $\lambda_\nu(\cdot)$ is bounded from above uniformly in ν,

$$\Pr(\mathcal{E}_k(T)) \geq \alpha(T) > 0 \tag{64}$$

for some positive function α. On the other hand,

$$\left\langle \omega = \mathbf{0} \;\middle|\; \mathcal{E}_k(T) \right\rangle_{\mu_{\nu,\lambda_\nu(\cdot)}(kT-\delta)} \geq \langle \omega = \mathbf{0} \rangle_{M_{\nu,\varepsilon}(kT-\delta)} . \tag{65}$$

Indeed, since $\int_s^{s+T} \lambda_\nu(t)\, dt < T(1-\varepsilon)$ for all s, the two processes $\mu_{\nu,\lambda_\nu(\cdot)}$ and $M_{\nu,\varepsilon}$ can be coupled in such a way that to each customer $C(t,\bar{\eta})$ of the process $\mu_{\nu,\lambda_\nu(\cdot)}$, who arrives at the moment t, $(l-1)T < t \leq lT$ and uses the server for time $\bar{\eta}$, it corresponds a customer $C'(\mathfrak{t}(t),\bar{\eta})$ of the process $M_{\nu,\varepsilon}$, who arrives at a later moment

$$\mathfrak{t}(t) = \left(\left\lfloor \frac{t}{T} \right\rfloor + 1 \right) T = lT$$

and needs the server for the same time duration $\bar{\eta}$. Hence the queue at every moment $t < kT$ of the process $M_{\nu,\varepsilon}$ is not shorter than the one for the process $\mu_{\nu,\lambda_\nu^{(k-1)T}(\cdot)}$, where

$$\lambda_\nu^{(k-1)T}(t) = \begin{cases} \lambda_\nu(t) & \text{if } t \leq (k-1)T, \\ 0 & \text{otherwice.} \end{cases} .$$

From (64), (65) we infer that for all $k \geq k_0$

$$\langle \omega = \mathbf{0} \rangle_{\mu_{\nu,\lambda_\nu(\cdot)}(kT-\delta)} \geq \frac{1}{2}\alpha(T)\, L(\varepsilon.\delta),$$

where k_0 is the smallest index k, for which $\langle \omega = \mathbf{0} \rangle_{M_{\nu,\varepsilon}(kT-\delta)} \geq \frac{1}{2} L(\varepsilon,\delta)$. (Of course, the value $k_0 = k_0(\nu)$ does depend on the initial state ν.)

Thus far we got the desired result (61) only for values $t \in [kT-\delta, kT]$, $k \geq k_0(\nu)$. To take care of the other values of t-s we should just make a change of variables and to start our process from all the different states $\mu_{\nu,\lambda_\nu(\cdot)}(t)$, $t \in [0,T]$. Be the convergence in (62) uniform in ν, that would be the end of the proof. However this is not the case. Yet, this does not create any problem, since the family of states $\left\{\mu_{\nu,\lambda_\nu(\cdot)}(t), t \in [0,T]\right\}$, which will be taken for initial states of the process $M_{*,\varepsilon}$, is compact, and on it the convergence in (62) is uniform indeed. ■

We finish this subsection with a statement about the regularity of the exit flow.

Lemma 14 *Let the function $p(t)$ satisfies the strong Lipschitz condition (8): for some C*

$$|p(t+\Delta t) - p(t)| \leq Cp(t)\,\Delta t. \tag{66}$$

Let ν be some initial state of GFP, and $\lambda(\cdot)$ be arbitrary rate function of arriving customers. (In particular one can take $\lambda = \lambda_\nu$, thus getting NMP.) Then the rate function $b(t)$ of the corresponding exit flow is Lipschitz, with Lipschitz constant independent of ν, λ.

Proof. Let t be fixed. The informal idea of the proof is the following: consider the elementary event ϖ, which contributes to the output rate $b(t)$, and let $\mathfrak{c}$ be the customer, who leaves the server at the moment t, being under the service for the time duration $\mathfrak{t}$. Then the elementary event ϖ', obtained from ϖ by enlarging the service time of $\mathfrak{c}$ from $\mathfrak{t}$ to $\mathfrak{t}+\Delta t$, contributes to $b(t+\Delta t)$. So we can get the desired result by comparing the probabilities of ϖ and ϖ'.

This correspondence, however, does not "cover" all the events, contributing to $b(t+\Delta t)$. Namely, the elementary events not covered by the above correspondence, are precisely those events ϖ'', for which the customer $\mathfrak{c}$, which is the last customer who has started his/her service before the moment t, is different from the customer $\mathfrak{c}''$, whose service terminates at $t+\Delta t$. This, though, means that the service time of $\mathfrak{c}''$ was less than Δt.

To implement this idea, let us write the measure of our process

$$\exp\left\{-I_\lambda(T)\right\}\sum_{n=0}^{\infty}\frac{1}{n!}\left[\prod_{i=1}^{n}\lambda(x_i)\,dx_i\prod_{i=1}^{n}p(l_i)\,dl_i\right]$$

on a segment $[0,T]$, with any $T > t+\Delta t$ (compare with (34)), as

$$d\Pi(\varpi) = \frac{1}{Z_T}\pi(\varpi)\,d\varpi,$$

with $Z_T = \exp\{I_\lambda(T)\}$ and $\pi(\varpi) = \prod_{i=1}^{n}\left[\lambda(x_i)\,p(l_i)\right]$, so

$$b(t) = \int_{B(t)} d\Pi(\varpi).$$

Here $B(t)$ is the manifold of all elementary events ϖ, which have a moment of service termination to happen at t. (We are treating here the case when

the initial state ν is concentrated on the configuration $\omega = \mathbf{0}$; the general case is totally similar.)

Let us split the rate

$$b(t+\Delta t) = \int_{B(t+\Delta t)} d\Pi(\varpi)$$

into two parts. The first one, $b'(t+\Delta t)$, is given by

$$b'(t+\Delta t) = \int_{B'(t+\Delta t)} d\Pi(\varpi),$$

where $B'(t+\Delta t) \subset B(t+\Delta t)$ is the image of the manifold $B(t)$ under the map $\varpi \rightsquigarrow \varpi'$, defined in the first paragraph of the present proof. Therefore

$$|b'(t+\Delta t) - b(t)| = \left| \int_{B(t)} \left(\frac{p(\mathfrak{t}+\Delta t)}{p(\mathfrak{t})} - 1 \right) d\Pi(\varpi) \right|,$$

where $\mathfrak{t} = \mathfrak{t}(\varpi)$ is the service time of the customer $\mathfrak{c}$ in ϖ, whose service terminates at the moment t. From (66) it follows that

$$|b'(t+\Delta t) - b(t)| \leq C\Delta t.$$

The remaining part $b''(t+\Delta t)$ of the rate $b(t+\Delta t)$,

$$b''(t+\Delta t) = \int_{B''(t+\Delta t)} d\Pi(\varpi),$$

corresponds to the event when some customer $\mathfrak{c}$ finishes his service during the time period $[t, t+\Delta t]$, while the next customer needs time at most Δt to be served. Such an event has probability below $b(t)(\Delta t)^2$, so the proof follows. ■

7.2 Estimates on the averaging kernels

Here we will estimate the densities $q_{\lambda,x}(t)$, entering into the relation $b(x) = [\lambda * q_{\lambda,x}](x)$.

Lemma 15 *The family $q_{\lambda,y}(t)$ is weakly continuous in y, for every λ. Also*

$$q_{\lambda,y}(t) \geq p(t) \Pr\{\textit{server is idle at the moment } y-t\}. \tag{67}$$

Lemma 16

$$q_{\lambda,y}(t) \leq \sum_{n=1}^{\infty} p^{*n}(t) \Pr\left\{N_t^{\lambda,y} \geq n-1\right\} \equiv \mathcal{Q}_{\lambda,y}(t), \qquad (68)$$

where $N_t^{\lambda,y}$ *is the random number of* λ*-Poisson points in the segment* $[y-t, y]$.

In particular, there exists a constant $\tilde{C} = \tilde{C}(p)$, *such that for* t *and all* λ, y

$$q_{\lambda,y}(t) \leq \tilde{C}. \qquad (69)$$

Proof. Both the relations (67) and (68) follow easily from the definition (28). To see (67) we note that, evidently, $c(u,t) \geq p(t)$. To get (68), we split the event $\mathcal{C}(u,t)$, entering (27), into the sum of events $\mathcal{C}_n(u,t)$, $n \geq 1$, where $\mathcal{C}_n(u,t)$ consists of all these outcomes when between the moment u of arrival of the first customer and the arrival of the customer who terminates during $[u+t, u+t+h]$ precisely $n-2$ other customers came. (The event $\mathcal{C}_1(u,t)$ consists from the outcomes when the first customer himself terminates during $[u+t, u+t+h]$.) But if $\mathcal{C}_n(u,t)$ holds, then two independent events have to happen:

$$\eta_1 + \ldots + \eta_n \in [t, t+h],$$

and

$$N_t^{\lambda,u} \geq n-1,$$

which imply (68).

To see (69), we use a rough form of (68):

$$q_{\lambda,y}(t) \leq \sum_{n=1}^{\infty} p^{*n}(t). \qquad (70)$$

Let $A = \sup_t p(t)$. Then it is immediate from (70) that for all $t \leq C$

$$q_{\lambda,y}(t) \leq A\left(1 + \sum_{n=1}^{\infty} \Pr\{\eta_1 + \ldots + \eta_n \leq C\}\right),$$

where η_i are i.i.d. random variables, distributed as η. But the probabilities $\Pr\{\eta_1 + \ldots + \eta_n \leq C\}$ decay exponentially in n, so the series converges for every t. It is a classical result of the renewal theory, that the sum (70) goes to a finite limit as $t \to \infty$, see e.g. the relation (1.17) of Chapter XI in [F]. That proves (69).

Continuity of $q_{\lambda,y}$ in y follows from the definition in a straightforward way. ■

We will need the compactness estimate on the distributions $q_{\lambda,y}(t)$. We will obtain them using the estimate (68). As the following statement shows, the estimate (68) is rather rough; we believe that all the moments of the distribution $q_{\lambda,y}(t)$ of order less than $1+\delta$ are finite.

Lemma 17 *Suppose that λ is such that for some T' and $\varepsilon' > 0$ and for all $T \geq T'$ and $s \geq 0$*

$$\int_s^{s+T} \lambda(t)\ dt < T(1-\varepsilon') \tag{71}$$

(see (55)). Then for any $b < \frac{\delta}{2}$

$$\int_0^{\infty} t^b q_{\lambda,y}(t)\ dt < C(\lambda, b) < \infty, \tag{72}$$

where $C(\lambda, b)$ depends on λ only via T' and ε'.

Proof of Lemma 17. We are going to use the simple estimate: for every random variable ζ and every $\varkappa > 0$

$$\tilde{Q}(T) \equiv \Pr\{\zeta > T\} \leq T^{-\varkappa}\mathbb{E}(|\zeta|^{\varkappa}). \tag{73}$$

We also will need an estimate on $\int_A^{\infty} t^a \tilde{q}(t)\ dt$, $a < \varkappa$, where $\tilde{q}$ is the density of ζ. We have:

$$\begin{aligned} \int_A^{\infty} t^a \tilde{q}(t)\ dt &= -\int_A^{\infty} t^a\, d\left(\tilde{Q}(t)\right) \\ &= A^a \tilde{Q}(A) + a\int_A^{\infty} t^{a-1}\tilde{Q}(t)\ dt. \end{aligned} \tag{74}$$

To apply (73) to (68) we will use the Dharmadhikari-Yogdeo estimate (see, e.g. [P], p.79): if ξ_i are independent centered random variables, then

$$\mathbb{E}\left(|\xi_1 + ... + \xi_n|^{2+\delta}\right) \leq R n^{\delta/2} \sum_1^n \mathbb{E}\left(|\xi_i|^{2+\delta}\right). \tag{75}$$

Here $R = R(\delta)$ is some universal constant.

Introducing $\xi_i = \eta_i - 1$ (see (13)), and using (73) with $\varkappa = 2 + \delta$ and (75) , we have (see (10))

$$\begin{aligned} Q_n(t) &\equiv \Pr\{\eta_1 + ... + \eta_n > t\} = \Pr\{\xi_1 + ... + \xi_n > t - n\} \\ &\leq RM_\delta (t-n)^{-(2+\delta)} n^{1+\delta/2}. \end{aligned} \tag{76}$$

To proceed, we use (68) to write

$$\begin{aligned} \int_0^\infty t^b q_{\lambda,y}(t)\, dt &\leq \int_0^\infty t^b \mathcal{Q}_{\lambda,y}(t)\, dt \\ &= \sum_{n=1}^\infty \left[\int_0^\infty t^b p^{*n}(t) \Pr\left\{N_t^{\lambda,y} \geq n-1\right\} dt\right]. \end{aligned} \tag{77}$$

Note that due to (71) there exists an $\alpha > 0$ such that $\mathbb{E}\left(N_t^{\lambda,y}\right) \leq (1-\alpha)\,t$ once t is large enough, uniformly in y. The first step is to estimate every summand by

$$\begin{aligned} &\int_0^\infty t^b p^{*n}(t) \Pr\left\{N_t^{\lambda,y} \geq n-1\right\} dt \\ &\leq \int_0^{n\left(1+\frac{\alpha}{2}\right)} t^b p^{*n}(t) \Pr\left\{N_t^{\lambda,y} \geq n-1\right\} dt + \int_{n\left(1+\frac{\alpha}{2}\right)}^\infty t^b p^{*n}(t)\, dt. \end{aligned} \tag{78}$$

Now, using (74) and (76) , we have for the second term in (78) :

$$\begin{aligned} \int_{n\left(1+\frac{\alpha}{2}\right)}^\infty t^b p^{*n}(t)\, dt &\leq \left[n\left(1+\frac{\alpha}{2}\right)\right]^b RM_\delta \left(\frac{\alpha}{2}n\right)^{-(2+\delta)} n^{1+\delta/2} \\ &\quad + bRM_\delta n^{1+\delta/2} \int_{n\left(1+\frac{\alpha}{2}\right)}^\infty t^{b-1}(t-n)^{-(2+\delta)}\, dt \\ &\leq Cn^{b-1-\delta/2}, \end{aligned}$$

where $C = C(\alpha, \delta, M_\delta)$.

The first term in (78) is negligible. To see that, we first observe:

Lemma 18 *Let* $0 < \nu < 1$, *and* N_n^ν *be a Poisson random variable:*

$$\Pr\{N_n^\nu = k\} = e^{-\nu n}\frac{(\nu n)^k}{k!}.$$

Then

$$\Pr\left\{N_n^\nu \geq n\right\} \leq \frac{1}{1-\nu} e^{-\frac{(1-\nu)^2}{2}n},$$

provided n is large enough.

Proof. Note first of all, that if $\chi > 0$ and $n > \chi$, then

$$e^{-\chi} \sum_{k \geq n} \frac{\chi^k}{k!} \leq e^{-\chi} \frac{\chi^n}{n!} \sum_{k \geq 0} \left(\frac{\chi}{n+1}\right)^k = e^{-\chi} \frac{\chi^n}{n!} \frac{1}{1 - \frac{\chi}{n+1}}.$$

In our case we thus have

$$\sum_{k \geq n} \Pr\left\{N_n^\nu = k\right\} \leq e^{-\nu n} \frac{(\nu n)^n}{n!} \frac{1}{1-\nu}.$$

By Stirling, for n large

$$\begin{aligned}\sum_{k \geq n} \Pr\left\{N_n^\nu = k\right\} &\leq \frac{1}{1-\nu} e^{-\nu n} \frac{\nu^n n^n}{n^n e^{-n}} \\ &= \frac{1}{1-\nu} e^{(1-\nu+\ln\nu)n} \\ &\leq \frac{1}{1-\nu} e^{-\frac{(1-\nu)^2}{2}n}.\end{aligned}$$

■

To estimate the integral $\int_0^{n\left(1+\frac{\alpha}{2}\right)} t^b p^{*n}(t) \Pr\left\{N_t^{\lambda,y} \geq n-1\right\} dt$ we note, that in the range of $t \in \left[0, n\left(1+\frac{\alpha}{2}\right)\right]$ we have

$$\Pr\left\{N_t^{\lambda,y} \geq n-1\right\} \leq \Pr\left\{N_{n\left(1+\frac{\alpha}{2}\right)}^{\lambda,y} \geq n-1\right\}.$$

We apply to the r.h.s. the last lemma, with $\nu = \frac{n-1}{n\left(1+\frac{\alpha}{2}\right)}$. Therefore, for all n large enough and uniformly in y

$$\begin{aligned}&\int_0^{n\left(1+\frac{\alpha}{2}\right)} t^b p^{*n}(t) \Pr\left\{N_t^{\lambda,y} \geq n-1\right\} dt \\ &\leq \frac{2}{\alpha} e^{-\frac{\alpha^2}{8}n} \int_0^{n\left(1+\frac{\alpha}{2}\right)} t^b p^{*n}(t)\, dt \\ &\leq \frac{2}{\alpha} e^{-\frac{\alpha^2}{8}n} \left[n\left(1+\frac{\alpha}{2}\right)\right]^b.\end{aligned}$$

Hence, the moment $\int_0^\infty t^b q_{\lambda,y}(t)\, dt$ is finite as soon as the series $\sum_n n^{b-1-\delta/2}$ converges, which happens when $b < \frac{\delta}{2}$. That proves Lemma 17. ■

8 The self-averaging relation: general case

Here we derive a formula, expressing the function $b(\cdot) = A(\mu, \lambda(\cdot))$ in terms of the functions $\lambda(\cdot)$, $p(\cdot)$ and the initial state μ of our non-stationary (GFP) Markov process. This will be the needed self-averaging relation (26). We remind the reader that μ is a probability measure on the set of pairs $\{(n,\tau)\}\cup$ $\mathbf{0}$.

Theorem 19 *Let $N(\mu) = q$, and the rate function $\lambda(\cdot)$ satisfies the conclusions of the Lemma 7:*

$$\int_s^{s+T} \lambda(t)\, dt < T(1-\varepsilon') \quad \textit{for all } T \geq T' > 0, \ \textit{all } s \geq 0 \textit{ and some } \varepsilon' > 0. \tag{79}$$

Then there exists the family of probability densities $q_{\lambda,\mu,x}(\cdot)$, $x > 0$, and the functionals $\varepsilon_{\lambda,\mu}(x)$ and $Q_{\lambda,\mu}(x)$, such that

$$b(x) = (1 - \varepsilon_{\lambda,\mu}(x))\left[\lambda * q_{\lambda,\mu,x}\right](x) + \varepsilon_{\lambda,\mu}(x)\, Q_{\lambda,\mu}(x). \tag{80}$$

Moreover,

$$\varepsilon_{\lambda,\mu}(x) \to 0 \ \textit{as } x \to \infty, \tag{81}$$

thought not necessarily uniformly in μ, while $Q_{\lambda,\mu}(x) \leq C$, uniformly in λ, μ and x.

Proof. We start by defining the functional $\varepsilon_{\lambda,\mu}(x)$. Note that the description of the realization of our process up to the moment x consists of the following data:

i) the initial configuration (n,τ), drawn from the distribution μ (with n to be the number of customers in the system before time zero, τ being the time the first one already spent in the server);

ii) the random set $0 < x_1 < ... < x_m < x$ (with random number m of points), which is a realization of the Poisson random field defined by the rate function λ (restricted to the segment $[0,x]$), independent of (n,τ) (the arrival moments of the customers, which come after time moment zero);

iii) one realization η_1 of the conditional random variable $\eta_\tau \equiv \left(\eta - \tau \,\middle|\, \eta > \tau\right)$ and $n+m-1$ independent realizations $\eta_k, k = 2,...,n+m$ of the random variable η (service times for the customers).

We denote by $\mathbb{P}_{\mu\otimes\lambda\otimes\eta}$ the corresponding (product) distribution. The difference $1-\varepsilon_{\lambda,\mu}(x)$ is by definition just the $\mathbb{P}_{\mu\otimes\lambda\otimes\eta}$-probability of the event

$$\sum_1^{n+m} \eta_k < x. \tag{82}$$

(If $n=0$, then by definition we put $\tau=0$; we put also $\sum_1^0 \equiv 0$.)

The meaning of the decomposition (80) can be explained now: the first term corresponds to the exit flow computed over those realizations where the relation (82) holds, while the second term represents the rest of the flow.

Let us prove (81) , that is that

$$\Pr\left\{\sum_1^{n+m} \eta_k > x\right\} \to 0 \,\text{as}\, x \to \infty.$$

To do this, we introduce two independent random variables:

$$S_\mu = \sum_1^n \eta_k,\ S_\lambda = \sum_{n+1}^{n+m} \eta_k.$$

Then for every $\alpha \in (0,1)$ we have

$$\begin{aligned}\Pr\left\{\sum_1^{n+m} \eta_k > x\right\} &= \Pr\left\{S_\mu + S_\lambda > x\right\} \\ &\le \Pr\left\{S_\mu > \alpha x\right\} + \Pr\left\{S_\lambda > (1-\alpha)\, x\right\}.\end{aligned}$$

Indeed, if $S_\mu + S_\lambda > x$, then either $S_\mu > \alpha x$, or else $S_\lambda > (1-\alpha)\, x$. Since S_μ is a random variable, the probability $\Pr\{S_\mu > \alpha x\}$ goes to zero for every α positive, as $x \to \infty$, though not necessarily uniformly in μ. For the second term we have

$$\begin{aligned}&\Pr\left\{S_\lambda > (1-\alpha)\, x\right\} \\ &= \sum_{m=1}^{\infty} \left(\int_{(1-\alpha)x}^{\infty} p^{*m}(t)\ dt\right) \Pr\left\{N^{\lambda,x} = m\right\}.\end{aligned}$$

Here $N^{\lambda,x}$ is the random number of points of the λ-Poisson field in $[0,x]$. Note that $\mathbb{E}\left(N^{\lambda,x}\right) < x\,(1-\varepsilon')$ once $x > T'$. Therefore we can apply the

same argument which was used in the proof of Lemma 17 when showing that the integral $\int_T^\infty \mathcal{Q}_{\lambda,y}(t) \to 0$ as $T \to \infty$, see (68). It implies that $\Pr\{S_\lambda > (1-\alpha)x\} \to 0$ once α is small enough, uniformly in λ, satisfying (79). That establishes (81).

Next we define the distributions $q_{\lambda,\mu,x}$. They are constructed from the random field of the rods $\{\eta_k, k=1,...,n+m\}$, defined above, placed at locations $\left\{\underbrace{0,...,0}_{n}, x_1,...,x_m\right\}$, via the procedure of resolution of conflicts, defined in the previous section. To do it we first introduce the rate $b_L(x)$ to be the exit rate of the conditional service process under the conditions that

$$\sum_1^n \eta_k = L, \quad \sum_{n+1}^{n+m} \eta_k < x - L. \tag{83}$$

We claim that for some probability distributions $q_{\lambda,L,x}$ we have

$$b_L(x) = [\lambda * q_{\lambda,L,x}](x).$$

The distribution $q_{\lambda,\mu,x}$ is then obtained by integration:

$$q_{\lambda,\mu,x} = \int q_{\lambda,L,x} \mathbb{P}_{\mu\otimes\lambda\otimes\eta}\left(\sum_1^n \eta_k \in dL\right).$$

(The random variable $\sum_1^n \eta_k$ is of course independent of the Poisson λ-field.) The output rate $b_L(x)$ corresponds to the situation when we have customers arriving at the moments $0, x_1,...,x_m$, which have serving times $L, \eta_{n+1},...,\eta_{n+m}$, and which satisfy the relation

$$L + \sum_{n+1}^{n+m} \eta_k < x.$$

So we have to repeat the construction of the Section 5 in the present situation. Few steps require some comments. The transition from the relation (37) to (38) uses the fact that for any s the measure $\prod_{i=1}^s p(l_i)\,dl_i$ is invariant under the coordinate permutations S_s in $\mathbb{R}^s$. But the same S_m symmetry evidently holds for the conditional distribution of the random vector $\left\{(\eta_k, k=n+1,...,n+m) \,\middle|\, \sum_{n+1}^{n+m} \eta_k < x - L\right\}$, since both the unconditional distribution and the distribution of the condition are S_m-invariant.

The next crucial step was the relation (41), stating that the functions $q_{\lambda,y}$ are probability distributions. It was based on the Theorem 4. The situation at hand is somewhat more delicate, since the rods we are dealing now with, are of two kinds: the first one has a non-random length L, produced by the initial state μ, while others are situated at the Poissonian locations $\{x_i\}$, defined by the rate function λ. However, under condition $\sum_{n+1}^{n+m} \eta_k < x - L$ the needed combinatorial statement (about the quantity $m!$) still holds, and is the content of the Theorem 6. These remarks allow one to carry over the construction of the Section 5, and so to establish the existence of the probability densities $q_{\lambda,L,x}$, and thus also $q_{\lambda,\mu,x}$. The upper and lower estimates on $q_{\lambda,\mu,x}$ are obtained in the same way as were the estimates for $q_{\lambda,x}$ in the preceding section.

The function $Q_{\lambda,\mu}(x)$ is the rate of exit flow of our process, conditioned by the event

$$\sum_{1}^{n+m} \eta_k \geq x.$$

The boundedness of the $Q_{\lambda,\mu}(x)$ follows from the following property of the service time distribution $p(x)$: for every $x, \tau,\ x > \tau > 0,\ 1 > t > 0$

$$\frac{p(x)}{p(x+t)} \leq C', \quad \frac{p\left(x-\tau \;\middle|\; \eta > \tau\right)}{p\left(x-\tau+t \;\middle|\; \eta > \tau\right)} \leq C'. \tag{84}$$

The relation (84) follows easily from the condition (8), with $C' = C'(C)$. To explain the boundedness, consider the elementary event

$$(n, \tau) \times \{x_1, ..., x_m : 0 < x_1 < ... < x_m < x\} \times \{\eta_1, ..., \eta_{n+m}\},$$

which contributes to the output flow inside the segment $[x, x + \Delta x]$, which flow is accounted by the second term of (80). That means that our rod configuration produces after resolution of conflicts a hit inside $[x, x + \Delta x]$, and also that

$$\sum_{1}^{n+m} \eta_k > x. \tag{85}$$

In the notation of the Section 6 it means that after resolution of conflicts the endpoint y_k of some (shifted) rod fits within $[x, x + \Delta x]$, for some $k \in$

$\{1,...,n+m\}$. Let $\bar{k}$ be the smallest such index. But then the elementary events

$$(n,\tau)\times\{x_1,...,x_m:0<x_1<...<x_m<x\}\times\{\eta_1,...,\eta_{\bar{k}-1},\eta_{\bar{k}}+t,\eta_{\bar{k}+1},...,\eta_{n+m}\},$$

with any $t\in(\Delta x,1)$, do not contribute to the output flow inside the segment $[x,x+\Delta x]$, while still satisfying (85). Therefore, due to (84), the probability that the customer would finish his service during the period $[x,x+\Delta x]$, is of the order of Δx, and, moreover,

$$Q_{\lambda,\mu}(x)\leq\frac{1}{C'}.$$

$\blacksquare$

Let now $\mathfrak{M}\in\mathcal{M}(\mathcal{M}_q(\Omega))$ be some invariant measure of the dynamical system (19). Then $\mathfrak{M}$-almost every state $\tilde{\mu}_0\in\mathcal{M}_q(\Omega)$ belongs to the family $\{\tilde{\mu}_t:-\infty<t<+\infty\}$, such that for all $\tau>0$, all t

$$\mathcal{T}_\tau(\tilde{\mu}_t)=\tilde{\mu}_{t+\tau}.$$

Let us fix one such family $\{\tilde{\mu}_t\}$. Then the function $\lambda(t)$, $-\infty<t<+\infty$, which for every $-\infty<\tau<+\infty$ satisfies on $[\tau,+\infty)$ the equation

$$\lambda(\cdot)=A(\tilde{\mu}_\tau,\lambda(\cdot),\tau),$$

is well defined. Then, according to the equation (80), for every τ, $-\infty<\tau<+\infty$, and for all $x\geq\tau$

$$\lambda(x)=(1-\varepsilon_{\lambda,\tilde{\mu}_\tau}(x))\left[\lambda*q_{\lambda,\tilde{\mu}_\tau,x}\right](x)+\varepsilon_{\lambda,\tilde{\mu}_\tau}(x)\,Q_{\lambda,\tilde{\mu}_\tau}(x). \qquad (86)$$

One would like to pass here to the limit $\tau\to-\infty$. According to (81), for every x we have $\varepsilon_{\lambda,\tilde{\mu}_\tau}(x)\to 0$ as $\tau\to-\infty$. Moreover, it is not difficult to show that in the same limit $q_{\lambda,\tilde{\mu}_\tau,x}(\cdot)\to q_{\lambda,x}(\cdot)$. So the following equation holds for λ:

$$\lambda(x)=\left[\lambda*q_{\lambda,x}\right](x),\quad -\infty<x<+\infty. \qquad (87)$$

By the methods developed below one can show that every bounded solution of (87) is a constant. Since, however, we are proving a stronger statement, that the dynamical system $\mathcal{T}_\tau$ has one fixed point on each $\mathcal{M}_q(\Omega)$, which is, moreover, globally attractive, we will not provide the details.

9 Self-averaging $\Longrightarrow$ relaxation: a warm-up

Before presenting the general proof that self-averaging implies relaxation, we consider the following simpler system: we have infinitely many servers, with service time η, distributed according to the probability density p. As the customer comes, he chooses any free server, and is served, leaving the system afterwards. The inflow is Poissonian, given by the rate function $f(x)$. If we impose the condition that the customers are coming at the rate they are living the system, we get the non-linear Markov process. The self-averaging relation (25) in such a case simplifies to

$$b(x) = [f * p](x).$$

Lemma 20 *Let $p(x)$ be some probability density with support, belonging to $\mathbb{R}^+$, satisfying the conditions of Section 2. Let f be a positive bounded function on $\mathbb{R}^1$. Suppose that*

$$f * p(x) = f(x) \text{ for all } x \geq 0. \tag{88}$$

Then $f(x) \to c$ as $x \to \infty$, for some $c > 0$.

Proof. Our statement follows easily from the well known results of the renewal theory. Let us introduce the function

$$\varphi(x) = \begin{cases} f(x) & \text{if } x < 0, \\ 0 & \text{if } x > 0. \end{cases}$$

Then the values of the function f for $x \geq 0$ can be recovered from the function φ and the fact that it satisfies (88). Indeed, iterating (88), we find:

$$f(x) = \begin{cases} [\varphi * p](x) + \left[\left(\varphi * p \Big|_{\{x \geq 0\}} \right) * \left(\sum_{n=1}^{\infty} p^{*n} \right) \right](x) & \text{for } x > 0, \\ \varphi(x) & \text{for } x < 0. \end{cases} \tag{89}$$

The function

$$s(x) = \sum_{n=1}^{\infty} p^{*n}(x)$$

is a key object of the renewal theory; in particular, it is known that it goes to a positive limit as $x \to \infty$; see for example the relation (1.17) of Chapter XI in [F]. From that and the relation (89) our claim follows. ■

In fact, the constant c can be computed: if m is the mean value of η, then

$$\lim_{x\to\infty} f(x) = \frac{1}{m}\int_0^{+\infty} [\varphi * p](x)\ dx.$$

This fact, as well as the renewal relation itself – $s(x) \to \mathit{const}$ as $x \to \infty$ – is a consequence of the statement that the probability density $p^{*n}(x)$ of the sum of i.i.d. random variables, $\eta_1 + ... + \eta_n$, is well approximated, due to the local limit theorem, by the Gaussian distribution

$$\frac{1}{\sqrt{2\pi n v}} e^{-(x-nm)^2/2nv},$$

where v is the variance of η. It becomes very flat as n increases.

10 Self-averaging $\Longrightarrow$ relaxation: probabilistic proof?

As was already said in Section 2, any function λ, defined for $x < 0$, and vanishing for $x < -T$, can be uniquely extended to $x \geq 0$ in such a way that the relation

$$A(\mathbf{0}, \lambda(\cdot), -T) = b(\cdot)$$

holds with $b(x) = \lambda(x)$ for $x \geq 0$. Therefore for every $x \geq 0$ we have

$$\lambda(x) = [\lambda * q_{\lambda,x}](x), \tag{90}$$

where $q_{\lambda,x}(\cdot)$ is a probability density supported by the semiaxis $\{y \geq 0\}$, and depending on λ via its restriction $\lambda \Big|_{\{y \leq x\}}$. Our goal is to show that (90) implies that $\lambda(x)$ relaxes to some constant c as $x \to \infty$.

Since the distributions $q_{\lambda,x}$ depend on $\lambda(\cdot)$ in a very complicated way, we have to treat a more general statement. Suppose a family of probability densities $q_x(\cdot)$, supported by the semiaxis $\{y \geq 0\}$, is given, where $x \geq 0$. Let $f(x)$ be a non-negative function, defined on $\mathbb{R}^1$, such that

$$\begin{aligned} & f(x) \leq C \text{ for } x < 0, \\ & f(x) = [f * q_x](x) \text{ for } x \geq 0. \end{aligned} \tag{91}$$

One would like to show that

$$\lim_{x\to\infty} f(x) = c, \tag{92}$$

for some $c \geq 0$. That will imply the relaxation needed.

Motivated by the analysis of the previous section, we will study the equation (91) by considering the corresponding non-stationary renewal process and the resulting inhomogeneous Markov random walk. Unfortunately, the relation (92) does not follow from (91) in general, and the reasons are probabilistic! Before explaining it let us "solve" (91).

So, let the family $\{q_x, x \geq 0\}$ be given; we solve (91) for f, given its restriction $f \Big|_{\{x<0\}}$. We do this in close analogy with the previous section, see (89). We put

$$f_0(x) = \begin{cases} f(x) & \text{for } x < 0 \\ 0 & \text{for } x \geq 0. \end{cases}$$

We define

$$f_{n+1}(x) = \begin{cases} f(x) & \text{for } x < 0, \\ [f_n * q_x](x) & \text{for } x \geq 0. \end{cases}$$

Then for every x the sequence $f_n(x)$ is increasing, and the function $f(x) = \lim_{n\to\infty} f_n(x)$ solves (91).

We can rewrite the function f in a different way. We define

$$g_1(x) = \begin{cases} [f_0 * q_x](x) & \text{for } x \geq 0, \\ 0 & \text{for } x < 0, \end{cases}$$

$$g_{n+1}(x) = [g_n * q_x](x). \tag{93}$$

Then for $x \geq 0$ we have

$$f(x) = \sum_{n\geq 1} g_n(x).$$

Now we will write the formula for g_n in terms of convolution. Here instead of the density p^{*n} of the sum S_n of i.i.d. random variables $\eta_1 + ... + \eta_n$ of the previous Section we have to consider the distribution $p_x^{(n)}$ of the inhomogeneous Markov walker $\bar{S}_{n,x}$, defined as follows. Remember that at each point $x \in \mathbb{R}^+$ we have the probability density q_x. So when our walker after some steps happen to arrive to the location x, then the next move is to the location $x + y$, with the increment $y > 0$ distributed with the density $q_x(y)$. The

random variable $\bar{S}_{n,x}$ is now defined as the position of the above described Markov walker after n steps, with the initial position x. With these notation we have, by (93):

$$g_{n+1}(x) = \left[g_1 * p_x^{(n)}\right](x).$$

Summarizing, we have for $x > 0$:

$$f(x) = g_1(x) + \sum_{n \geq 2} \left[g_1 * p_x^{(n)}\right](x),$$

in analogy with (89). As in the previous section, the Local Limit Theorem for the Markov chain $\bar{S}_{n,x}$ would imply the relaxation (92).

We have to note, however, that the relation between the validity of the LLT for our Markov chain and the validity of the relation (92) is more complicated. Namely, LLT for $\bar{S}_{n,x}$ might fail, while the relaxation (92) might still remain valid – or else fail as well! First of all, let us explain that even the Central Limit Theorem for $\bar{S}_{n,x}$ might not hold, notwithstanding the family $q_x(\cdot)$ to have very nice compactness properties. To give one example, consider the family of probability densities $u_x(t)$, $x \in \mathbb{R}^1$, where all $u_x(\cdot)$ have for their support the segment $[0,1]$, and satisfy there $0 < c < u_x(t) < C < \infty$, uniformly in x and t. We define now

$$q_x(t) = u_x(t - (1 - \{x\})),$$

where $\{\cdot\}$ stays for the fractional part. Then all $q_x(\cdot)$-s have their supports within the segment $[0,2]$. But the random variables $\bar{S}_{n,x}$ do not have CLT behavior! Indeed, the random variable $\bar{S}_{n,x}$ is localized in the segment $[\lfloor x \rfloor + n, \lfloor x \rfloor + n + 1]$, where $\lfloor \cdot \rfloor$ denotes the integer part. So the variance of $\bar{S}_{n,x}$ remains bounded in n ! Nevertheless, for this example it can be shown that the relation (92) still holds, and that involves certain statement of the type of Perron-Frobenius theorem for our Markov chain. Further modification of this example, when

$$q_x(t) = u_x(t - (2 - \{x\})),$$

results in the Markov chain with two classes, and in this case both the CLT and the relation (92) might fail.

We conjecture here that the LLT theorem for the sums $\bar{S}_{n,x}$ holds, if the family $q_x(\cdot)$ of transition densities has the following additional property:

- For some k, K, $0 < k < K < \infty$,

$$k \le \frac{q_{x_1}(t)}{q_{x_2}(t)} \le K, \tag{94}$$

provided at least one of the values $q_{x_i}(t)$ is positive.

The condition (94) is reminiscent of the *positivity of ergodicity coefficient* condition, introduced by Dobrushin [D1] in his study of the limit theorems for the non-stationary Markov chains.

In what follows we will take another road, and we get the relaxation property by analytic methods, which seems in our case to be simpler. But we still use probability theory, though not the CLT. It would be interesting to obtain the desired result by proving the corresponding limit theorem.

11 Self-averaging $\Longrightarrow$ relaxation: finite range case

In this section we prove the relaxation for the solution of the equation (91) in the finite range case.

The reader of the paper of course understands that for our initial problem we have to consider the infinite range case. We think nevertheless that the finite range case is of independent interest, and moreover, it holds in a much more general setting than the infinite-range case. This is why we devote to it the present Section. Its content is not used in what follows.

Theorem 21 *Suppose that*

$$0 \le f(x) \le C \text{ and is continuous for } x < 0,$$
$$f(x) = [f * q_x](x) \text{ for } x \ge 0,$$

while the following three conditions on the family q_x hold:

i) for some T and all x

$$\int_0^T q_x(t)\, dt = 1, \quad q_x(t) = 0 \text{ for } t > T,$$

ii) the family $q_x(\cdot)$ depends on x continuously,

iii) for $0 \leq t \leq T$

$$C \geq q_x(t) \geq \kappa(t) > 0, \tag{95}$$

with continuous positive $\kappa(t)$ *.*

Then the limit exists:

$$\lim_{x \to \infty} f(x) = c \geq 0.$$

The property (95) holds for the NMP, as follows from the relations (67), (61) and (69).

Proof. *i*) We know that the function f is continuous and bounded, $0 \leq f \leq C$. So if there exists a value X such that f is monotone for $x \geq X$, then the function f has to be constant for $x \geq X + T$, and we are done. So we are left with the case when the function f has infinitely many points of local maxima and local minima, which go to ∞.

ii) Given a local maximum, x_0, we will construct now a sequence x_i of local maximums, $i = 0, -1, -2, ..., -n = -n(f, x_0)$ such that

- $x_0 > x_{-1} > x_{-2} > ...,$
- $x_i - x_{i-1} < 2T, \;\; x_i - x_{i-2} \geq T$ for all i,
- $0 < x_{-n} < 2T$,
- $f(x_{i-1}) \geq f(x)$ for any $x_{i-1} \leq x$, and $f(x_{i-1}) > f(x_i)$,
- for every $x \in [x_{i-1}, x_i - T]$ we have $f(x) \geq f(x_i)$ (of course if the segment is non-empty, i.e. $x_i - x_{i-1} > T$).

The construction is the following. Let x_0 be some point of local maxima. Since

$$f(x_0) = \int_0^T f(x_0 - t)\, q_{x_0}(t)\; dt,$$

we have $f(x_0) < F(x_0) \equiv \sup\{f(x) : x \in [x_0 - T, x_0]\}$, unless f is a constant on $[x_0 - T, x_0]$, in which case we are done. Let $y = \inf\{x \in [x_0 - T, x_0] : f(x) = F(x_0)\}$. If $y > x_0 - T$, or if $y = x_0 - T$ and is a local maximum, we define $x_{-1} = y$. In the opposite case we have that the point $x_0 - T$ is not a local maximum of the function f on the segment $[x_0 - 2T, x_0 - T]$. We then consider two cases.

In the first one we suppose that the function f on the segment $[x_0 - 2T, x_0 - T]$ takes values below $\bar{F} = \frac{f(x_0)+f(x_0-T)}{2}$. Let $[y, x_0 - T] \subset [x_0 - 2T, x_0 - T]$ be the largest segment for which the inequality $f(x) \geq \bar{F}$ holds for every $x \in [y, x_0 - T]$. We define x_{-1} to be the leftmost point of maximum of f in $[y, x_0 - T]$.

In the opposite case we consider the set $S = \{x \in [x_0 - 2T, x_0 - T] : f(x) \geq f(x_0 - T)\}$. It contains other points besides $x_0 - T$. However, it can not contain all the segment $[x_0 - 2T, x_0 - T]$. Since f is not a constant on $[x_0 - 2T, x_0 - T]$, $\sup_S f > f(x_0 - T)$. Let $z \in (x_0 - 2T, x_0 - T)$ be such that $f(z) < f(x_0 - T)$. Let $S_1 = S \cap [z, x_0 - T]$. We necessarily have that $\sup_{S_1} f > f(x_0 - T)$ as well. We define x_{-1} to be any point in S_1 where $f(x_{-1}) = \sup_{S_1} f$. Clearly, x_{-1} is a local maxima of f, while $x_0 - x_{-1} < 2T$.

We proceed to define the sequences x_i by induction, $i = 0, -1, -2, \ldots$. It is not excluded that $x_i \in [x_{i+1} - T, x_{i+1}]$ for some i. However that means in particular that the point x_i is the first maximum point of the function f on the segment $[x_{i+1} - T, x_{i+1}]$, and since $f(x_{i-1}) > f(x_i)$, we have that $x_{i-1} \notin [x_{i+1} - T, x_{i+1}]$ for all i. We stop when we arrive to a first value below $2T$.

iii) In the same way, starting from a local minima y_0, we can construct a sequence y_i of local minima, such that

- $y_0 > y_{-1} > y_{-2} > \ldots,$
- $y_i - y_{i-1} < 2T, \;\; y_i - y_{i-2} \geq T$ for all i,
- $0 < y_{-n} < 2T$,
- $f(y_{i-1}) \leq f(x)$ for any $y_{i-1} \leq x$, and $f(y_{i-1}) < f(y_i)$,
- for every $x \in [y_{i-1}, y_i - T]$ we have $f(x) \leq f(y_i)$ (if the segment is non-empty).

We can suppose additionally that $x_0 \geq y_0 \geq x_{-1}$.

iv) Note that the (finite) sequence x_i do depend on the initial local minima x_0, which was used for the starter. The bigger x_0 is, the longer the sequence x_i is. So let us introduce the sequence $x_0^{(N)}$ of such starters, and we suppose that $x_0^{(N)} \geq N$. In that way we will obtain the family $x_i^{(N)}$ of sequences of local maximums of f, $i = 0, -1, \ldots, -n\left(f, x_0^{(N)}\right)$, with $n\left(f, x_0^{(N)}\right) \to \infty$ as $N \to$

∞. (Of course, in well may happen that for different N-s the corresponding sequences share many common terms.)

Denote by M the limit $\liminf_{N\to\infty} f\left(x_0^{(N)}\right)$. In the same way we can introduce the limit $m = \limsup_{N\to\infty} f\left(y_0^{(N)}\right)$. Clearly, $M \geq m$, and if we can show that $M = m$, then we are done. So we suppose that $M - m > 0$, and we will bring that to contradiction.

v) Let us fix $\varepsilon > 0$, $\varepsilon < \frac{M-m}{10}$, which is possible if $M - m > 0$. Then one can choose N so large, that at least 99% of terms of the sequence $f\left(x_{i-1}^{(N)}\right) - f\left(x_i^{(N)}\right)$ are less than $\frac{\varepsilon^2}{2}$. We will fix that value of N, and we will omit N from our notation. Therefore without loss of generality we can assume that for some i (in fact, for many) we have $f(x) < f(x_i) + \varepsilon^2$ for all $x \in [x_i - T, x_i]$. Therefore for the set $K \equiv \{x \in [x_i - T, x_i] : f(x) > f(x_i) - \varepsilon\}$ we have:

$$\int_{K-(x_i-T)} q_{x_i}(t)\ dt > 1 - \varepsilon. \tag{96}$$

Hence, for its Lebesgue measure we have

$$\operatorname{mes}\{K\} \geq \frac{1-\varepsilon}{C},$$

with $C = \sup f$.

Consider now the corresponding sequence of minima, $\{y_k\}$, and the segments $[y_k - T, y_k]$. We claim that the set K has to belong to the union of these segments. That would be evident if the segments in question were covering the corresponding region without any holes. However, that is not necessarily the case, and there can be holes between the segments, since in general the differences $y_i - y_{i-1}$ can be bigger than T. Yet, this does not present a problem, since by construction the function f is smaller than m outside the union of the segments $[y_k - T, y_k]$, which implies that the set K indeed is covered by these segments. Since diam $(K) \leq T$, there exists $k = k(K)$, such that $K \subset [y_{k-1} - T, y_{k-1}] \cup [y_k - T, y_k] \cup [y_{k+1} - T, y_{k+1}]$. Without loss of generality we can assume the set K "fits into $[y_k - T, y_k]$", in the sense that

$$\operatorname{mes}\{K \cap [y_k - T, y_k]\} \geq \frac{\operatorname{mes}\{K\}}{3} \geq \frac{1-\varepsilon}{3C},$$

while we have $f(x) > f(y_k) - \varepsilon^2$ for all $x \in [y_k - T, y_k]$. So we have

$$\int_{\{K \cap [y_k - T, y_k]\} - (y_k - T)} q_{y_k}(t)\ dt \geq \bar{\kappa}\left(\frac{1-\varepsilon}{3C}\right), \tag{97}$$

where we define the function $\bar{\kappa}(\alpha)$ by

$$\bar{\kappa}(\alpha) = \inf_{A \subset [0,T]:\,\mathrm{mes}\{A\} \geq \alpha} \int_A \kappa(t)\ dt.$$

By construction, the set $K \cap [y_k - T, y_k]$ is disjoint from the set $L \subset [y_k - T, y_k]$, which is defined by $L = \{x \in [y_k - T, y_k] : f(x) < f(y_k) + \varepsilon\}$. Since

$$f(y_k) = \int_0^T f(y_k - t)\, q_{y_k}(t)\ dt,$$

we have similar to (96) that

$$\int_{L - (y_k - T)} q_{y_k}(t)\ dt > 1 - \varepsilon. \tag{98}$$

But since $q_{y_k}(t)\ dt$ is a probability measure, we should have that

$$\bar{\kappa}\left(\frac{1-\varepsilon}{3C}\right) + 1 - \varepsilon \leq 1,$$

because of (97), (98). This, however, fails once ε is small enough. ■

12 Self-averaging $\Longrightarrow$ relaxation: infinite range case

We return to the equation (91), $f(x) = [f * q_x](x)$. Now we will not suppose that the measures q_x have finite support. Instead we suppose that

1. The family q_x has the following compactness property: for every $\varepsilon > 0$ there exists a value $K(\varepsilon)$, such that

   $$\int_0^{K(\varepsilon)} q_x(t)\ dt \geq 1 - \varepsilon \tag{99}$$

   uniformly in x.

2. For every T the (monotone continuous) function

$$F_T(\delta) = \inf_{x \geq X(T)} \inf_{\substack{D \subset [0,T]: \\ \text{mes} D \geq \delta}} \int_D q_x(t)\ dt \tag{100}$$

is positive once $\delta > 0$, for some choice of the function $X(T) < \infty$.

3. The family q_x is such that the function f, with solves (91), is Lipschitz, with Lipschitz constant $\mathcal{L} = \mathcal{L}(\{q_\cdot\})$.

As we know from the Section 7, these conditions are indeed satisfied in the specific case of the non-linear Markov process and the equation (90). Indeed, (99) follows from Lemma 7 and Lemma 17, (100) – from Lemma 15 and Lemma 13, while Lipschitz property follows from Lemma 14.

Theorem 22 *Let f satisfies $f(x) = [f * q_x](x)$ for $x \geq 0$, with the kernels q_x having three properties listed above. Then f relaxes to a constant value as $x \to \infty$.*

12.1 Approaching stationary point

Lemma 23 *i) Let $M = \limsup_{x\to\infty} f(x)$. Then for every T and every ε given there exists some value K_1, such that*

$$\inf_{x \in [K_1, K_1+T]} f(x) \geq M - \varepsilon.$$

ii) Let $m = \liminf_{x\to\infty} f(x)$. Then for every T and every ε given there exists some value K_2, such that

$$\sup_{x \in [K_2, K_2+T]} f(x) \leq m + \varepsilon.$$

Moreover, the conclusions of the lemma remains valid if the function f satisfies a weaker equation (see (80))

$$f(x) = (1 - \varepsilon(x))[f * q_x](x) + \varepsilon(x) Q(x), \tag{101}$$

with $\varepsilon(x) \to 0$ as $x \to \infty$ and $Q(\cdot) \leq C$.

Proof. *i)* Let $\delta > 0$. Then there exists a value $S = S(\delta) > 0$, such that for all $x > S$ we have $f(x) < M + \delta$, and $\varepsilon(x) Q(x) < \frac{\delta}{2}$. Further, there exists a value $R = R(\delta) > S$, such that for all $y \geq R$

$$\int_{R-S}^{\infty} q_y(t)\, dt < \delta,$$

see (99). Finally, there exists a point $y = y(\delta) > R + T$, such that $f(y) > M - \frac{\delta}{2}$. Due to the equation (101) we have

$$f(y) = (1 - \varepsilon(y)) \left[\int_0^{y-S} f(y-t)\, q_y(t)\ dt + \int_{y-S}^{\infty} f(y-t)\, q_y(t)\ dt \right] + \varepsilon(y)\, Q(y).$$

Let $\Delta > 0$, and $A = \{x \in [y-T, y] : f(x) < M - \Delta\}$, while $a = \int_A q_y(t)\ dt$. We want to show that the measure a has to be small for small δ. Splitting the first integral into two, according to whether the point $y - t$ is in A or not, we have

$$M - \delta < a(M - \Delta) + (1 - a - \delta)(M + \delta) + \delta C,$$

so

$$a < \delta \frac{C + 2 - M}{\Delta},$$

which goes to zero with δ, provided Δ is fixed. Therefore

$$\text{mes}\{A\} \leq F_T^{-1}\left(\delta \frac{C + 2 - M}{\Delta}\right),$$

(see (100)). Since $F_T^{-1}(u) \to 0$ as $u \to 0$, that proves that for any given Δ the Lebesgue measure $\text{mes}\{A\} \to 0$ as $\delta \to 0$. Since the function f is Lipschitz, we conclude that $\inf_{x \in [y-T, y]} f(x) \geq M - \Delta - \mathcal{L}\text{mes}\{A\} \geq M - 2\Delta$, provided δ is small enough. Taking $\Delta = \varepsilon/2$ finishes the proof.

ii) Let $\delta > 0$. Then there exists a value $S > 0$, such that for all $x > S$ we have $f(x) > m - \delta$. Again, take $R > S$, such that for all $y \geq R$

$$\int_0^{R-S} q_y(t)\, dt > 1 - \delta.$$

Finally, there exists a point $y > R + T$, such that $f(y) < m + \delta$. Due to the equation (101) we have

$$f(y) > (1 - \varkappa) \left[\int_0^{y-S} f(y-t)\, q_y(t)\ dt + \int_{y-S}^{\infty} f(y-t)\, q_y(t)\ dt \right], \quad (102)$$

where $\varkappa$ can be supposed arbitrarily small. Let $\Delta > 0$, and $A = \{t \in [0,T] : f(y-t) > m + \Delta\}$, while $a = \int_A q_y(t)\ dt$. We want to show that the measure a has to be small for small δ. Splitting the first integral into two, according to whether the point $y-t$ is in A or not, and disregarding the second one, we have

$$m + \delta > (1-\varkappa)\left[a(m+\Delta) + (1-\delta-a)(m-\delta)\right].$$

For $\varkappa$ so small that $\varkappa\left[a(m+\Delta) + (1-a-\delta)(m-\delta)\right] < \delta$, we have

$$m + 2\delta > a(m+\Delta) + (1-a-\delta)(m-\delta),$$

so

$$a < \delta\frac{m+3}{\Delta},$$

which goes to zero with δ, provided Δ is fixed. Therefore

$$\operatorname{mes}\{A\} \leq F_T^{-1}\left(\delta\frac{m+3}{\Delta}\right),$$

and the rest of the argument coincides with that of the part i). ■

12.2 Absorbing by stationary point

We now will show that if f satisfies (91), then the property $\inf_{x\in[K,K+T]} f(x) \geq M - \varepsilon$ implies that for all $x > K + T$

$$f(x) > M - \varepsilon - c(T), \tag{103}$$

with $c(T) \to 0$ as $T \to \infty$. That clearly implies relaxation. (Note that we do not claim that (103) holds for the solutions of (101)). We will show it under the extra assumption that the distribution $p(\cdot)$ has finite moment of some order above 4. This assumption will be used only throughout the rest of the present subsection.

Using the linearity of (91), we will rewrite our problem slightly, in order to simplify the notation.

Let the function $f \geq 0$ satisfies $f(x) = [f * q_x](x)$ for $x > 0$, and

i) $f(x) > 1$ for $x \in [-T, 0]$,

ii) for some $\beta > 1$ and $B < \infty$ and for every x we have

$$\int_0^\infty t^\beta q_x(t)\ dx \leq B, \tag{104}$$

compare with (72). We want to derive from that data that for some $c(T)>0$, $c(T)\to 0$ as $T\to\infty$

$$f(x)>1-c(T) \text{ for all } x>0.$$

Denote by

$$g_0(x)=\begin{cases} 1 & x\in[-T,0] \\ 0 & x\notin[-T,0] \end{cases}.$$

Since $f\geq g$, we have $f(x)\geq g_1(x)=[g_0*q_x](x)$ for $x\geq 0$. We define $g_1(x)=g_0(x)$ for $x<0$. Iterating, we have $f(x)\geq g_n(x)$, where

$$g_n(x)=\begin{cases} g_0(x) & x<0 \\ [g_{n-1}*q_x](x) & x\geq 0 \end{cases}.$$

Hence, $f(x)\geq g_\infty(x)$. The function $g_\infty(x)$ has the following probabilistic interpretation: we have a Markov chain on $\mathbb{R}^1$, where transition from the point x is governed by transition densities q_x to make the step (to the left), (and which steps to the left are defined in an arbitrary way for $x\leq 0$); then the value $g_\infty(x)$ for $x>0$ is the probability that starting from x we will visit the interval $[-T,0]$. The question now is about the lower bound on $g_\infty(x)$ over all possible q_x from our class.

So let us take $x>0$, and let start the Markov chain X_n from x, (i.e. $X_0=x$), which goes to the left, and which makes a transition from y to $y-t$ with the probability $q_y(t)\,dt$. We need to know the probability of the event

$$\mathbb{P}_x\{\text{there exists } n \text{ such that } X_n\in[-T,0]\}.$$

In other words, we want to know the probability of $X_{\{\cdot\}}$ visiting $[-T,0]$. We would like to show that

$$\mathbb{P}_x\left\{X_{\{\cdot\}} \text{ visits } [-T,0]\right\}\geq\gamma(\beta,B,T) \tag{105}$$

with

$$\gamma(\beta,B,T)\to 1 \text{ as } T\to\infty$$

uniformly over the families q_x from our class.

Note, however, that in general such an estimate does not hold. For example, the process $X_{\{\cdot\}}$ can well stay positive for all times. The more interesting example where the process goes to $-\infty$, follows, so we will need further restrictions on the family q_x.

Example. Let T be given. We will construct the family q_x^T from our class (104), such that for the corresponding process $X_{\{\cdot\}}^T$

$$\mathbb{P}_x\left\{X_{\{\cdot\}}^T \text{visits } [0,T]\right\} = 0.$$

We define $q_x^T(t)$ for $x \in (k, k+1]$ with integer $k \neq 0$ to be any distribution localized in the segment $[k-1, k]$ (the uniform distribution on $[k-1, k]$ is OK). For $x \in (\frac{1}{2^k}, \frac{1}{2^{k-1}}]$, $k = 1, 2, ...$, it is defined by

$$q_x^T(t) = \begin{cases} e^{-t} & \text{if } t > T+1 \\ 2^{k+1}\left(1 - \int_{T+1}^{\infty} e^{-t} dt\right) & \text{if } t \in \left[x - \frac{1}{2^k}, x - \frac{1}{2^{k+1}}\right] \\ 0 & \text{otherwice.} \end{cases}$$

For $x \leq 0$ it is defined in an arbitrary way. ■

The mechanism of violating the relation (105) is that the time the process $X_{\{\cdot\}}^T$ can spend in the segment $[0, 1]$ is unbounded in T. As the following theorem shows, this feature is the only obstruction for the statement desired to hold.

Theorem 24 *Consider the Markov chain $X_{\{\cdot\}}$ defined above via the transition densities $q_x(t)$. Suppose that condition (104) holds, and that in addition these densities are uniformly bounded in the vicinity of the origin: for all real x and all t in the segment $[0, 1]$, say,*

$$q_x(t) \leq C. \tag{106}$$

Then for some $\gamma(\beta, B, C, T) \to 1$ as $T \to \infty$ we have:

$$\mathbb{P}_x\left\{X_{\{\cdot\}} \textit{visits } [-T, 0]\right\} \geq \gamma(\beta, B, C, T).$$

The condition (106) holds in the case of NMP, see estimate (69) from Lemma 16.

Proof. We will estimate the probability of the complementary event:

$$\begin{aligned} &\mathbb{P}_x\left\{X_{\{\cdot\}} \text{misses } [-T, 0]\right\} \\ &= \sum_{k=0}^{\infty} \int_0^x \left[\int_{y+T}^{\infty} q_y(t)\ dt\right] P_k(x, dy). \end{aligned}$$

Here $P_k(x, dy)$ is the probability distribution of the chain $X_{\{\cdot\}}$ after k steps, and the expression $\left[\int_{y+T}^{\infty} q_y(t)\ dt\right] P_k(x, dy)$ is the probability that the chain $X_{\{\cdot\}}$ arrives after k steps to the location y, and then makes a jump over the segment $[-T, 0]$. (In our case the probability of the event that $X_{\cdot}$ never becomes negative equals zero.) So we have

$$\begin{aligned}
&\mathbb{P}_x\left\{X_{\{\cdot\}} \text{ misses } [-T, 0]\right\} \\
&\leq \int_0^x B(y+T)^{-\beta} \sum_{k=0}^{\infty} P_k(x, dy) \\
&\leq \sum_{n=0}^{[x]+1} B(n+T)^{-\beta} \sum_{k=0}^{\infty} P_k(x, [n, n+1]),
\end{aligned}$$

where $P_k(x, [n, n+1])$ is the probability of the event $X_k \in [n, n+1]$, and where in the second line we are using the following simple estimate:

$$\int_r^{\infty} q_y(t)\ dt = r^{-\beta} \int_r^{\infty} r^{\beta} q_y(t)\ dt \leq r^{-\beta} \int_0^{\infty} t^{\beta} q_y(t)\ dt.$$

Now,

$$\begin{aligned}
&\sum_{k=0}^{\infty} \mathbb{P}_x\{X_k \in [n, n+1]\} \qquad (107)\\
&= \sum_{k=0}^{\infty} \sum_{l<k} \mathbb{P}_x\{X_k \in [n, n+1], X_l > n+1, X_{l+1} \in [n, n+1]\} \\
&= \sum_{l=0}^{\infty} \mathbb{P}_x\{X_l > n+1, X_{l+1} \in [n, n+1]\} \\
&\times \sum_{k>0} \mathbb{P}_x\left\{X_{l+k} \in [n, n+1] \,\middle|\, X_l > n+1, X_{l+1} \in [n, n+1]\right\}.
\end{aligned}$$

Let now the random variables ζ_i be i.i.d., uniformly distributed in the segment $\left[0, \frac{1}{C}\right]$, where C is the same as in (105). Then is easy to see that

$$\mathbb{P}_x\left\{X_{l+k} \in [n, n+1] \,\middle|\, X_l > n+1, X_{l+1} \in [n, n+1]\right\} \leq \Pr\{\zeta_1 + \ldots + \zeta_k \leq 1\}.$$

Since the last probability decays exponentially in k, while

$\sum_{l=0}^{\infty} \mathbb{P}_x \{X_l > n+1, X_{l+1} \in [n, n+1]\} = 1$, we conclude that

$$\sum_{k=0}^{\infty} \mathbb{P}_x \{X_k \in [n, n+1]\} \leq K(C).$$

Since the series $\sum n^{-\beta}$ converges for $\beta > 1$, the proof follows. ■

13 Self-averaging $\Longrightarrow$ relaxation: noisy case

In this section we establish the relaxation for the NMP with general initial condition. Using the fact that the Poisson rate $\lambda(x) = \lambda_\mu(x)$ of the NMP with initial state μ satisfies the equation

$$\lambda(x) = (1 - \varepsilon_{\lambda,\mu}(x)) [\lambda * q_{\lambda,\mu,x}](x) + \varepsilon_{\lambda,\mu}(x) Q_{\lambda,\mu}(x),$$

we will prove the following

Theorem 25 *Let the initial state μ of the NMP μ_t is such that both the expected service time $S(\mu)$ and the mean queue length $N(\mu)$ are finite. Then the limit $c = \lim_{x\to\infty} \lambda(x)$ exists; moreover, $\mu_t \to \nu_c$ as $t \to \infty$, where ν_c is the invariant measure of NMP, such that $\lambda_{\nu_c}(x) \equiv c$. Also $N(\nu_c) = N(\mu)$.*

We are not able to prove this theorem in the generality of the previous Sections. Below we will use all the specific features of the NMP-s, and in particular we will use the comparison between different NMP-s and GFP-s, corresponding to various initial states and input rates. The comparison mentioned is based on the coupling arguments.

13.1 Coupling

Definition 26 *Let μ_1, μ_2 be two states on Ω. We call the state μ_1 to be **higher** than μ_2, $\mu_1 \succcurlyeq \mu_2$, if there exists a coupling $P[d\omega_1, d\omega_2]$ between the states μ_1, μ_2, with the property:*

$$P\left[(\Omega \times \Omega)^{>}\right] = 1,$$

where

$$(\Omega \times \Omega)^{>} = \{[(n_1, \tau_1), (n_2, \tau_2)] \in \Omega \times \Omega : n_1 \geq n_2\}.$$

Clearly, if $\mu_1 \succcurlyeq \mu_2$, then $N(\mu_1) \geq N(\mu_2)$.

Next, we introduce the stronger relation.

Definition 27 *Let μ_1, μ_2 be two states on Ω. We call the state μ_1 to be **taller** than μ_2, $\mu_1 \succeq \mu_2$, if there exists a coupling $P[d\omega_1, d\omega_2]$ between the states μ_1, μ_2, with the property:*

$$P\left[(\Omega \times \Omega)^{\gg}\right] = 1,$$

where

$$(\Omega \times \Omega)^{\gg} = \{[(n_1, \tau_1), (n_2, \tau_2)] \in \Omega \times \Omega : \tau_1 = \tau_2, n_1 \geq n_2 \text{ or } (n_2, \tau_2) = \mathbf{0}\}.$$

Lemma 28 *Let $\mu_1(0)$, $\mu_2(0)$ be two initial states on Ω at $t = 0$, and $\lambda_1(t)$, $\lambda_2(t)$, $t \geq 0$ be two Poisson densities of the input flows. The service time distribution is the same η as before. Let $\mu_i(t)$ be two corresponding GFP-s. Suppose that $\mu_1(0) \succeq \mu_2(0)$, and that $\lambda_1(t) \geq \lambda_2(t)$. Then $\mu_1(t) \succcurlyeq \mu_2(t)$, so in particular*

$$N(\mu_1(t)) \geq N(\mu_2(t)).$$

Also, there exists a coupling between the processes such that for almost every trajectory $(\omega_1(t), \omega_2(t))$

$$S(\omega_1(t)) \geq S(\omega_2(t)). \tag{108}$$

Proof. To see this let us construct the coupling between the processes $\mu_i(t)$. Let us color the customers, arriving according to the $\lambda_2(t)$ flow, as red. We also assign the red color to the customers which were present at time $t = 0$ from the initial state $\mu_2(0)$. Let $\gamma(t) = \lambda_1(t) - \lambda_2(t)$, and consider $\gamma(t)$ as the extra input flow of blue customers (with independent service times). We also add blue customers at time $t = 0$, which are needed to complete the state $\mu_2(0)$ up to $\mu_1(0)$. Then the total (color blind) flow coincides with λ_1 flow, while the total (color blind) process coincides with $\mu_1(t)$.

The service rule for the two-colored process is color blind: all the customers are served in order of their arrival time. We claim now that along every coupled trajectory $(\omega_1(t), \omega_2(t))$ we have $r(\omega_1(t)) \geq r(\omega_2(t))$, where $r(\cdot)$ is the number of red customers at the moment t, waiting to be served. That evidently will prove our statement.

Clearly, the number $r(\omega(t))$ is the difference,

$$r(\omega(t)) = \mathcal{A}(\omega(t)) - \mathcal{B}(\omega(t)),$$

where $\mathcal{A}(\omega(t))$ is the total number of red customers, having arrived before t, while $\mathcal{B}(\omega(t))$ is the total number of red customers, who left the system before t. Clearly, $\mathcal{A}(\omega_1(t)) = \mathcal{A}(\omega_2(t))$. Let us show that $\mathcal{B}(\omega_1(t)) \leq \mathcal{B}(\omega_2(t))$.

This is easy to see once one visualizes the procedure of resolving the rod conflicts, which corresponds to our service rule, for the two-colored rod case. Namely, one has first to put all the red rods, and resolve all their conflicts by shifting some of them to the right accordingly. The number of thus obtained rods to the left of the point t is the number $\mathcal{B}(\omega_2(t))$. Clearly, if one adds some blue rods to the red ones, then each red rod would be shifted to the right by at least the same amount as without the blue rods. As a result, every red rod would either stay where it was, or move to the right, so indeed $\mathcal{B}(\omega_1(t)) \leq \mathcal{B}(\omega_2(t))$.

The relation (108) is evident. ■

13.2 Convergence

Consider a General Flow Process $\mu(t)$ with initial state $\mu(0) = \bar{\nu}$ at $T = 0$ and the input rate $\lambda(t) \equiv c < 1$ (i.e. the usual queueing system $M|GI|1$). This is an ergodic process, so the weak limit

$$\lim_{t\to\infty} \mu_{\bar{\nu},c}(t) = \nu_c$$

exists and does not depend on the initial state $\bar{\nu}$. We would like to show that if $N(\bar{\nu}) < \infty$, then also

$$\lim_{t\to\infty} N(\mu_{\bar{\nu},c}(t)) = N(\nu_c) \tag{109}$$

(see (18)). This however is not true in general, and extra assumptions are needed in order to have such convergence.

Lemma 29 *Suppose additionally that $S(\bar{\nu}) < \infty$. Then*

$$\lim_{t\to\infty} N(\mu_{\bar{\nu},c}(t)) = N(\nu_c).$$

Moreover, for every $s < \infty$ the convergence $N(\mu_{\bar{\nu},c}(t)) \to N(\nu_c)$ is uniform on the set of all initial states $\bar{\nu}$ satisfying $S(\bar{\nu}) \leq s$.

Proof of Lemma 29. Since $\lim_{t\to\infty}\mu_{\bar{\nu},c}(t)=\nu_c$, $\lim_{t\to\infty}N(\mu_{\bar{\nu},c}(t))\geq N(\nu_c)$. To prove the equality we need to show the uniform integrability for the family of random variables $N_{\mu_{\bar{\nu},c}(t)}$, which means the following property: for every $\varkappa>0$ there exists a value $s_\varkappa$ such that for all t

$$\mathbb{E}_{\mu_{\bar{\nu},c}(t)}\left(N(\omega)\,\mathbf{I}_{N(\omega)\geq s_\varkappa}\right)<\varkappa,$$

where $\mathbf{I}$ stands for the indicator.

Note first that it is enough to show the uniform integrability of the family $S_{\mu_{\bar{\nu},c}(t)}$ of random variables. Indeed, consider the event $N(\omega)\geq N$. Then the conditional $\mu_{\bar{\nu},c}(t)$-probability of the event $S(\omega)\geq\frac{1}{2}N$ under the condition $N(\omega)\geq N$ goes to 1 as $N\to\infty$, since $\mathbb{E}(\eta)=1$. Therefore if the family $S_{\mu_{\bar{\nu},c}(t)}$ is not uniformly integrable, so is $N_{\mu_{\bar{\nu},c}(t)}$.

To get uniform integrability we prove the following stochastic domination:

Lemma 30

$$S_{\mu_{\bar{\nu},c}(t)}\preccurlyeq S_{\bar{\nu}}+S_{\nu_c}. \tag{110}$$

(Here in the rhs we mean the sum of two independent random variables.)

Since the expectation $S(\bar{\nu})$ of the first random variable is finite by assumption, while the expectation of the second equals $N(\nu_c)$ and so is also finite, the uniform integrability of the family $S_{\mu_{\bar{\nu},c}(t)}$ follows.

Proof of Lemma 30. To prove (110) we will use the following construction. Let $x_1,\eta_1;x_2,\eta_2;...$ be a realization of the flow of customers. It means that at the moment x_1 the first customer comes, which needs the time η_1 to be served, at the moment x_2 the second comes, etc. To every such realization we can assign the function $W(x)$, which is the remaining time duration needed for the server to serve all the customers who came before the moment x. That is,

$$W(x)=\begin{cases}0 & \text{for } x<x_1,\\ \max\{\eta_1-(x-x_1),0\} & \text{for } x_1\leq x<x_2,\\ \max\{W(x_2-0)+\eta_2-(x-x_2),0\} & \text{for } x_2\leq x<x_3, \text{ etc.}\end{cases}$$

If μ_t is a process of states of our server with $\mu_0=\delta_{\mathbf{0}}$, and $\{x_1,\eta_1;x_2,\eta_2;...\}$ is its realization, then the random variable $W(t)$ is the same as S_{μ_t}. With obvious modification the W function is defined for a process with non-empty initial state μ_0.

Consider now two processes: $\mu_{\nu_c,c}(t)$ and $\mu_{\delta_{\mathbf{0}},c}(t)$. The first one is stationary. Since evidently $\nu_c \succcurlyeq \delta_{\mathbf{0}}$, we have $\mu_{\nu_c,c}(t) \succcurlyeq \mu_{\delta_{\mathbf{0}},c}(t)$, i.e. $\nu_c \succcurlyeq \mu_{\delta_{\mathbf{0}},c}(t)$ for all t. It means also that we can couple the two processes in such a way that $W^{\mu_{\nu_c,c}}(t) \geq W^{\mu_{\delta_{\mathbf{0}},c}}(t)$ with probability one.

Let us see now how the two processes – $W^{\mu_{\delta_{\mathbf{0}},c}}(t)$ and $W^{\mu_{\bar{\nu},c}}(t)$ – are related. We consider the natural coupling between $\mu_{\delta_{\mathbf{0}},c}$ and $\mu_{\bar{\nu},c}$, where the latter process is obtained by adding to a general configuration $\{x_1,\eta_1;x_2,\eta_2;...\}$ of the former one extra customer $\{x_0,\Pi_0\}$, with $x_0=0$ and the independent random variable Π_0 distributed according to $S_{\bar{\nu}}$. The trajectory $W^{\mu_{\bar{\nu},c}}(t)$ is obtained in the following way: one considers first the function

$$\tilde{W}(x) = \Pi_0 - x + \sum_{i\geq 1} \eta_i \chi_{x_i}(x),$$

where

$$\chi_a(x) = \begin{cases} 1 & \text{if } x \geq a, \\ 0 & \text{otherwice.} \end{cases}$$

Let x_0 be the first moment when $\tilde{W}(x)$ vanishes. Then

$$W^{\mu_{\bar{\nu},c}}(x) = \begin{cases} \tilde{W}(x) & \text{if } x \leq x_0, \\ W^{\mu_{\delta_{\mathbf{0}},c}}(x) & \text{otherwice.} \end{cases}$$

From this the relation (110) follows immediately.

The uniformity of convergence follows from the fact that the function S is a compact function on Ω, once the function $R_\eta(\tau)$ is unbounded (see (11)). (This compactness means that for every s the set $\{\omega\in\Omega : S(\omega)\leq s\}$ is compact.) As a result, the family of initial states $\bar{\nu}$ satisfying $S(\bar{\nu})\leq s$ is weakly compact as well, which together with continuity of the function $N(\mu_{\bar{\nu},c}(t))$ in $\bar{\nu}$ and t provides the claim needed. If the function $R_\eta(\tau)$ is uniformly bounded in τ, then for some ξ and C the exponential moment $\mathbb{E}(\exp\{\xi\eta_\tau\}) \leq C$ for all τ. Therefore the family of all possible probability distributions F_θ of the form

$$F_\theta(x) = \int F_{\eta_\tau}(x)\ d\theta(\tau),$$

where θ runs over all probability measures on the semiaxis $\{\tau\geq 0\}$, is compact. That again implies the uniformity. ■

■

13.3 End of the proof in noisy case

Let $\mu_{\nu,\lambda_\nu(\cdot)}(t)$ be the non-linear Markov process with the initial state ν, having finite mean queue, $N(\nu) < \infty$, and finite expected service time $S(\nu)$. We will show that the function $\lambda(t) \equiv \lambda_\nu(t)$ goes to a limit as $t \to \infty$. The idea is the following:

Suppose $m = \liminf_{t\to\infty} \lambda(t) < \limsup_{t\to\infty} \lambda(t) = M$. As we already know from Lemma 23, for every T, K and every $\varepsilon > 0$ there exist values $K_1, K_2 > K$ such that

$$\sup_{x\in[K_1,K_1+T]} \lambda(x) \leq m + \varepsilon, \tag{111}$$

while

$$\inf_{x\in[K_2,K_2+T]} \lambda(x) \geq M - \varepsilon. \tag{112}$$

We want to bring this to contradiction, arguing as follows:

- First of all, we note that the mean queue, $N\left(\mu_{\nu,\lambda_\nu(\cdot)}(t)\right)$ does not change in time, staying equal to the initial value $N(\nu)$.

  On the other hand:

- We can compare the state $\mu_{\nu,\lambda_\nu(\cdot)}(K_1+T)$ with the state $\mu_{\left[\mu_{\nu,\lambda_\nu(\cdot)}(K_1)\right],m+\varepsilon}(T)$. Due to (111), the latter is higher, so

$$N\left(\mu_{\left[\mu_{\nu,\lambda_\nu(\cdot)}(K_1)\right],m+\varepsilon}(T)\right) \geq N(\nu). \tag{113}$$

  By the same reasoning,

$$N\left(\mu_{\left[\mu_{\nu,\lambda_\nu(\cdot)}(K_2)\right],M-\varepsilon}(T)\right) \leq N(\nu). \tag{114}$$

- We then claim that once T is large enough, the state $\mu_{\left[\mu_{\nu,\lambda_\nu(\cdot)}(K_1)\right],m+\varepsilon}(T)$ is close to the equilibrium $\nu_{m+\varepsilon}$, so in particular

$$N\left(\mu_{\left[\mu_{\nu,\lambda_\nu(\cdot)}(K_1)\right],m+\varepsilon}(T)\right) \leq N(\nu_{m+\varepsilon}) + \varepsilon', \tag{115}$$

  with $\varepsilon' = \varepsilon'(T) \to 0$ as $T \to \infty$. By the same reasoning,

$$N\left(\mu_{\left[\mu_{\nu,\lambda_\nu(\cdot)}(K_2)\right],M-\varepsilon}(T)\right) \geq N(\nu_{M-\varepsilon}) - \varepsilon''. \tag{116}$$

- Since $N(\nu_{M-\varepsilon}) > N(\nu_{m+\varepsilon})$ once ε is small, the choice of ε' and ε'' such that $\varepsilon' + \varepsilon'' < N(\nu_{M-\varepsilon}) - N(\nu_{m+\varepsilon})$ makes it possible to deduce from the relations (113)-(116) that

$$N(\nu) \geq N(\nu_{M-\varepsilon}) - \varepsilon'' > N(\nu_{m+\varepsilon}) + \varepsilon' \geq N(\nu),$$

  which is inconsistent with the properties of the relation $>$ between the real numbers.

We need to prove the relations (115) and (116). It turns out that the relation (116) is easier. Indeed, to show it, we can compare the state $\mu_{\left[\mu_{\nu,\lambda_\nu(\cdot)}(K_2)\right],M-\varepsilon}(T)$ with the state $\mu_{\mathbf{0},M-\varepsilon}(T)$. The latter is evidently lower :

$$N\left(\mu_{\left[\mu_{\nu,\lambda_\nu(\cdot)}(K_2)\right],M-\varepsilon}(T)\right) \geq N(\mu_{\mathbf{0},M-\varepsilon}(T)).$$

Since $\mu_{\mathbf{0},M-\varepsilon}(T)$ is also lower than $\nu_{M-\varepsilon}$,

$$N(\mu_{\mathbf{0},M-\varepsilon}(T)) \leq N(\nu_{M-\varepsilon}). \tag{117}$$

Since $\mu_{\mathbf{0},M-\varepsilon}(T) \to \nu_{M-\varepsilon}$ as $T \to \infty$, (117) implies that $N(\mu_{\mathbf{0},M-\varepsilon}(T)) \to N(\nu_{M-\varepsilon})$, which proves (116).

In the above proof the important step was to replace the state $\mu_{\nu,\lambda_\nu(\cdot)}(K_2)$ with a lower state $\mathbf{0}$, which is in fact the lowest. Turning to (115), we see that this step can not be mimicked there, since there is no highest state! So, to proceed, we need some apriori upper bound on the state $\mu_{\nu,\lambda_\nu(\cdot)}(K_1)$.

Lemma 31 *Let ν be an arbitrary initial state, with $N(\nu) < \infty$. Then there exist $\bar{c}(\nu) < 1$ and $\mathcal{T} < \infty$, such that for every $t > \mathcal{T}$*

$$\lambda_\nu(t) < \bar{c}(\nu).$$

Proof. The statement of the lemma is equivalent to the fact that $M = \limsup_{t\to\infty} \lambda_\nu(t) < 1$. So suppose the opposite, that $M \geq 1$. As we then know from Lemma 23, for every $\mathcal{T}$ and every $\varepsilon > 0$ we can find a segment $[K, K+\mathcal{T}]$, such that $\lambda_\nu(t) > 1-\varepsilon$ for all $t \in [K, K+\mathcal{T}]$. This, however, contradicts to the statement (55) of Lemma 7. ■

Since $S(\nu) < \infty$, and the rate $\lambda_\nu(t)$ is uniformly bounded, the function $S(\mu_{\nu,\lambda_\nu}(t))$ is finite for every t. It can grow, but after the moment $\mathcal{T}$, obtained in the last Lemma, it stays bounded from above by the value $S(\mu_{\nu,\lambda_\nu}(\mathcal{T}))+$

$S(\nu_{\bar{c}})$, because of Lemma 30. So without loss of generality we can assume that the initial state ν itself is such that $N(\nu) < \infty$, $S(\nu) < \infty$, while $\lambda_\nu(t) < \bar{c} < 1$ for all $t > 0$ and so $S(\mu_{\nu,\lambda_\nu}(t)) \leq S(\nu) + S(\nu_{\bar{c}})$. Therefore the family of states $\{\mu_{\nu,\lambda_\nu}(t), t \geq 0\}$ is weakly compact. Hence for every ε' there exists T such that for all t

$$N\left(\mu_{\mu_{\left[\nu,\lambda_\nu(\cdot)(t)\right]},m+\varepsilon}(T)\right) \leq N(\nu_{m+\varepsilon}) + \varepsilon'.$$

This proves (115).

14 Conclusions

In this paper we have proven the Poisson Hypothesis for the information networks, for the case of the "mean-field" model of the network. We have found the domain of its validity, and we will show in the forthcoming paper [RS] that beyond this domain Poisson Hypothesis is violated even in the mean-field case, and the dependence of the initial condition does not vanish with time. We strongly believe that the methods we have developed here – in particular, the self-averaging relation – are relevant not only for mean-field models, but also for more realistic ones.

The following problems appear as the natural continuation of our work.

- The study of Poisson Hypothesis for the case of the service times forming an ergodic random process, rather than the sequence of i.i.d. random variables.
- The study of PH for the case of customers of several identities, with service times depending on their identity.
- The study of PH for more general graphs.

We are going to work on these problems in the near future.

Acknowledgment. *We would like to thank our colleagues – in particular, Yu. Golubev, T. Liggett, O. Ogievetsky, G. Olshansky, S. Pirogov, A. Vladimirov – for valuable discussions and remarks, concerning this paper. We are grateful to our referee for his remarks, which resulted in a substantial improvement of the presentation of the paper. We are grateful also to*

the Institute for Pure and Applied Mathematics (IPAM) at UCLA, for the uplifting atmosphere and support during the Spring 2002 program on Large Scale Communication Networks, where part of this work was done.

References

[B] Borovkov, A. A. Talk at INRIA, 1995, see also [Bu].

[Bu] Borovkov, K. A. Propagation of chaos in queueing networks. Teor. Veroyatnost. i Primenen. 42 (1997), no. 3, 449-460; translation in Theory Probab. Appl. 42 (1997), no. 3, 385-394 (1998).

[DF] Delcoigne, F.; Fayolle, G. Thermodynamical limit and propagation of chaos in polling systems. Markov Process. Related Fields 5 (1999), no. 1, 89–124.

[D1] Dobrushin, R.L. Central limit theorem for inhomogeneous Markov chains. Probability Theory and its Applications, I, II, v. 1, p. 72-89, v. 1, p. 365-425, 1956.

[D2] Dobrushin, R.L. Private communications, see also [DS].

[DKV] Dobrushin, R.L., Karpelevich, F.I., and Vvedenskaya, N.D."Queueing Systems with Choice of Shortest Queue - Asymptotic Approach", Problemy Peredachi Informatsii", 1996, V.32, N.1, pp.20-36

[DS] R.L. Dobrushin and Suhov, Yu. M. Asymptotic investigation of star-shape message switching networks with a large number of radial rays. Problems of information transmission, 12 (1976), 49-65.

[F] Feller W. An introduction to Probability Theory and Its Applications, 2-nd edition, v. 2. John Wiley & Sons, NY, 1971.

[KR1] Karpelevich, F. I.; Rybko, A. N. Asymptotic behavior of a symmetric closed queueing network at a thermodynamic limit. (Russian) Problemy Peredachi Informatsii 36 (2000), no. 2, 69–95; translation in Probl. Inf. Transm. 36 (2000), no. 2, 154–179

[KR2] Karpelevich, F. I.; Rybko, A. N. Thermodynamic limit for the mean field model of simple symmetrical closed queueing network. Markov Process. Related Fields 6 (2000), no. 1, 89–105.

[K] Kleinrock, L., Communication Nets; Stochastic Message Flow and Delay, McGraw-Hill Book Company, New York, 1964.

[L] Liggett, Thomas M. Interacting particle systems. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], 276. Springer-Verlag, New York, 1985.

[M1] McKean, H. P., Jr. A class of Markov processes associated with nonlinear parabolic equations. Proc. Nat. Acad. Sci. U.S.A., **56,** 1966, 1907–1911.

[M2] McKean, H. P., Jr. An exponential formula for solving Boltzmann's equation for a Maxwellian gas. J. Combinatorial Theory, **2,** 1967, 358–382.

[P] Petrov, V. V. Sums of independent random variables. Translated from the Russian by A. A. Brown. Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 82. Springer-Verlag, New York-Heidelberg, 1975.

[RS] Rybko, A. N.; Shlosman, S.B.: Poisson Hypothesis for Information Networks II. Violation and Phase Transitions, in preparation.

[St] Stolyar, A. L. The asymptotics of stationary distribution for a closed queueing system. (Russian) Problemy Peredachi Informatsii 25 (1989), no. 4, 80–92; translation in Problems Inform. Transmission 25 (1989), no. 4, 321–331 (1990)

[S] Stoyan, D. Qualitative Eigenschaften und Abschatzungen stochastischer Modelle, Akademie-Verlag, Berlin, 1977.